



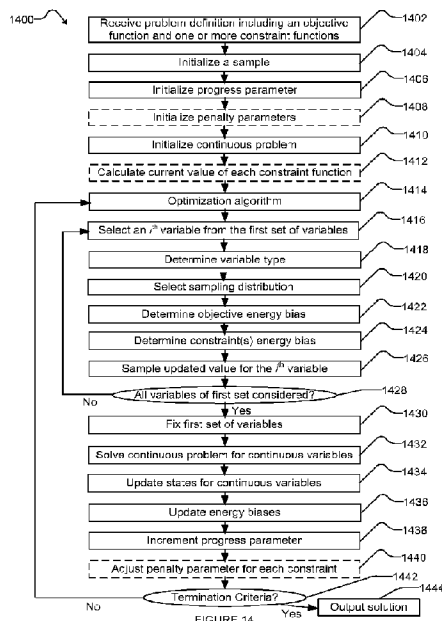
(12) **DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION**

(13) **A1**

(86) **Date de dépôt PCT/PCT Filing Date:** 2022/03/30
(87) **Date publication PCT/PCT Publication Date:** 2022/10/20
(85) **Entrée phase nationale/National Entry:** 2023/10/11
(86) **N° demande PCT/PCT Application No.:** IB 2022/000201
(87) **N° publication PCT/PCT Publication No.:** 2022/219399
(30) **Priorités/Priorities:** 2021/04/13 (US63/174,097);
2021/09/30 (US63/250,466)

(51) **Cl.Int./Int.Cl. G06N 10/60** (2022.01),
G06N 10/20 (2022.01), **G06N 10/80** (2022.01)
(71) **Demandeur/Applicant:**
D-WAVE SYSTEMS INC., CA
(72) **Inventeurs/Inventors:**
SADEGHI ESFAHANI, HOSSEIN, CA;
RAHMANI, MOHSEN, CA;
ZUCCA, ALEX, CA;
BERNOUDY, WILLIAM W., CA
(74) **Agent:** LAMBERT INTELLECTUAL PROPERTY LAW

(54) **Titre : SYSTEMES ET PROCEDES D'AMELIORATION D'EFFICACITE DE CALCUL INFORMATIQUE DE DISPOSITIFS A BASE DE PROCESSEUR LORS D'UNE RESOLUTION DE MODELES QUADRATIQUES CONTRAINTS**
(54) **Title: SYSTEMS AND METHODS FOR IMPROVING COMPUTATIONAL EFFICIENCY OF PROCESSOR-BASED DEVICES IN SOLVING CONSTRAINED QUADRATIC MODELS**



(57) **Abrégé/Abstract:**

Systems and methods for optimization algorithms, updating samples, and penalizing constraint violations are discussed. A method for updating samples includes receiving a problem definition with an objective function and constraint functions, an initial sample, and a value for a progress parameter. For each variable a total energy change is determined based on an objective energy change based on the sample value for the variable and one or more terms of the objective function that include the variable and a constraint energy change based on the sample value for the variable and each of the constraint functions defined by the variable. A sampling distribution is selected based on the variable type and an updated value is sampled based on the total energy change and the progress parameter. An updated sample is returned with an updated value for each variable of the set of variables. Such may improve operation of processor-based systems.

Date Submitted: 2023/10/11

CA App. No.: 3215170

Abstract:

Systems and methods for optimization algorithms, updating samples, and penalizing constraint violations are discussed. A method for updating samples includes receiving a problem definition with an objective function and constraint functions, an initial sample, and a value for a progress parameter. For each variable a total energy change is determined based on an objective energy change based on the sample value for the variable and one or more terms of the objective function that include the variable and a constraint energy change based on the sample value for the variable and each of the constraint functions defined by the variable. A sampling distribution is selected based on the variable type and an updated value is sampled based on the total energy change and the progress parameter. An updated sample is returned with an updated value for each variable of the set of variables. Such may improve operation of processor-based systems.

SYSTEMS AND METHODS FOR IMPROVING COMPUTATIONAL EFFICIENCY OF PROCESSOR-
BASED DEVICES IN SOLVING CONSTRAINED QUADRATIC MODELS

FIELD

This disclosure generally relates to systems and methods for improving
5 efficiency of processor-based devices in solving constrained quadratic models, for instance
employing a hybrid approach that takes advantage of digital processors and analog processors.

BACKGROUND

Quantum Devices

Quantum devices are structures in which quantum mechanical effects are
10 observable. Quantum devices include circuits in which current transport is dominated by
quantum mechanical effects. Such devices include spintronics and superconducting circuits.
Both spin and superconductivity are quantum mechanical phenomena. Quantum devices can
be used for measurement instruments, in computing machinery, and the like.

15 Quantum Computation

A quantum computer is a system that makes direct use of at least one
quantum-mechanical phenomenon, such as superposition, tunneling, and entanglement, to
perform operations on data. The elements of a quantum computer are qubits. Quantum
computers can provide speedup for certain classes of computational problems such as
20 computational problems simulating quantum physics.

Quantum Processor

A quantum processor may take the form of a superconducting quantum
processor. A superconducting quantum processor may include a number of superconducting
25 qubits and associated local bias devices. A superconducting quantum processor may also
include coupling devices (also known as couplers or qubit couplers) that selectively provide
communicative coupling between qubits.

A quantum processor is any computer processor that is designed to leverage at least one quantum mechanical phenomenon (such as superposition, entanglement, tunneling, etc.) in the processing of quantum information. Regardless of the specific hardware implementation, all quantum processors encode and manipulate quantum information in quantum mechanical objects or devices called quantum bits, or “qubits;” all quantum processors employ structures or devices for communicating information between qubits; and all quantum processors employ structures or devices for reading out a state of at least one qubit. A quantum processor may include a large number (e.g., hundreds, thousands, millions, etc.) of programmable elements, including but not limited to: qubits, couplers, readout devices, latching devices (e.g., quantum flux parametron latching circuits), shift registers, digital-to-analog converters, and/or demultiplexer trees, as well as programmable sub-components of these elements such as programmable sub-components for correcting device variations (e.g., inductance tuners, capacitance tuners, etc.), programmable sub-components for compensating unwanted signal drift, and so on.

Further details and embodiments of exemplary quantum processors that may be used in conjunction with the present systems and devices are described in, for example, U.S. Patent Nos. 7,533,068; 8,008,942; 8,195,596; 8,190,548; and 8,421,053.

Hybrid Computing System Comprising a Quantum Processor

A hybrid computing system can include a digital computer communicatively coupled to an analog computer. In some implementations, the analog computer is a quantum computer, and the digital computer is a classical computer.

The digital computer can include a digital processor that can be used to perform classical digital processing tasks described in the present systems and methods. The digital computer can include at least one system memory which can be used to store various sets of computer- or processor-readable instructions, application programs and/or data.

The quantum computer can include a quantum processor that includes programmable elements such as qubits, couplers, and other devices. The qubits can be read out via a readout system, and the results communicated to the digital computer. The qubits

and the couplers can be controlled by a qubit control system and a coupler control system, respectively. In some implementations, the qubit and the coupler control systems can be used to implement quantum annealing on the analog computer.

5 Quantum Annealing

Quantum annealing is a computational method that may be used to find a low-energy state of a system, typically preferably the ground state of the system. The method relies on the underlying principle that natural systems tend towards lower energy states, as lower energy states are more stable. Quantum annealing may use quantum effects, such as
10 quantum tunneling, as a source of delocalization to reach an energy minimum.

A quantum processor may be designed to perform quantum annealing and/or adiabatic quantum computation. An evolution Hamiltonian can be constructed that is proportional to the sum of a first term proportional to a problem Hamiltonian and a second term proportional to a delocalization Hamiltonian, as follows:

$$15 \quad H_E \propto A(t)H_P + B(t)H_D$$

where H_E is the evolution Hamiltonian, H_P is the problem Hamiltonian, H_D is the delocalization Hamiltonian, and $A(t)$, $B(t)$ are coefficients that can control the rate of evolution, and typically lie in the range [0,1].

In some implementations, a time varying envelope function can be placed on
20 the problem Hamiltonian. A suitable delocalization Hamiltonian is given by:

$$H_D \propto -\frac{1}{2} \sum_{i=1}^N \Delta_i \sigma_i^x$$

where N represents the number of qubits, σ_i^x is the Pauli x-matrix for the i^{th} qubit and Δ_i is the single qubit tunnel splitting induced in the i^{th} qubit. Here, the σ_i^x terms are examples of “off-diagonal” terms.

25 A common problem Hamiltonian includes a first component proportional to diagonal single qubit terms and a second component proportional to diagonal multi-qubit terms, and may be of the following form:

$$H_P \propto -\frac{\epsilon}{2} [\sum_{i=1}^N h_i \sigma_i^z + \sum_{j>i}^N J_{ij} \sigma_i^z \sigma_j^z]$$

where N represents the number of qubits, σ_i^z is the Pauli z-matrix for the i^{th} qubit, h_i and J_{ij} are dimensionless local fields for the qubits, and couplings between qubits, and ϵ is some characteristic energy scale for H_P .

5 Here, the σ_i^z and $\sigma_i^z \sigma_j^z$ terms are examples of “diagonal” terms. The former is a single qubit term and the latter a two-qubit term.

Throughout this specification, the terms “problem Hamiltonian” and “final Hamiltonian” are used interchangeably unless the context dictates otherwise. Certain states of the quantum processor are energetically preferred, or simply preferred, by the problem
10 Hamiltonian. These include the ground states and may include excited states.

Hamiltonians such as H_D and H_P in the above two equations may be physically realized in a variety of different ways. A particular example is realized by an implementation of superconducting qubits.

15 Sampling

Throughout this specification and the appended claims, the terms “sample”, “sampling”, “sampling device”, and “sample generator” are used.

In statistics, a sample is a subset of a population, i.e., a selection of data taken from a statistical population. In electrical engineering and related disciplines, sampling relates
20 to taking a set of measurements of an analog signal or some other physical system.

In many fields, including simulations of physical systems, and computing, especially analog computing, the foregoing meanings may merge. For example, a hybrid computer can draw samples from an analog computer. The analog computer, as a provider of samples, is an example of a sample generator. The analog computer can be operated to
25 provide samples from a selected probability distribution, the probability distribution assigning a respective probability of being sampled to each data point in the population. The population can correspond to all possible states of the processor, and each sample can correspond to a respective state of the processor.

Markov Chain Monte Carlo

Markov Chain Monte Carlo (MCMC) is a class of computational techniques which include, for example, simulated annealing, parallel tempering, population annealing, and other techniques. A Markov chain may be used, for example when a probability
5 distribution cannot be used. A Markov chain may be described as a sequence of discrete random variables, and/or as a random process where at each time increment the state only depends on the previous state. When the chain is long enough, aggregate properties of the chain, such as the mean, can match aggregate properties of a target distribution.

The Markov chain can be obtained by proposing a new point according to a
10 Markovian proposal process (generally referred to as an “update operation”). The new point is either accepted or rejected. If the new point is rejected, a new proposal is made, and so on. New points that are accepted are ones that make for a probabilistic convergence to the target distribution. Convergence is guaranteed if the proposal and acceptance criteria satisfy detailed balance conditions, and the proposal satisfies the ergodicity requirement. Further,
15 the acceptance of a proposal can be done such that the Markov chain is reversible, i.e., the product of transition rates over a closed loop of states in the chain is the same in either direction. A reversible Markov chain is also referred to as having detailed balance. Typically, in many cases, the new point is local to the previous point.

The foregoing examples of the related art and limitations related thereto are
20 intended to be illustrative and not exclusive. Other limitations of the related art will become apparent to those of skill in the art upon a reading of the specification and a study of the drawings.

BRIEF SUMMARY

Improving the computational efficiency and/or accuracy of the operation of
25 processor-based devices is generally desirable and is particularly desirable in the case of using processor-based devices as solves to solve constrained quadratic models.

One method of addressing constraints when solving an optimization problem is to convert the constraints to part of the objective function using slack variables, and then

minimize over the new objective function. However, in some implementations, this may result in a high computational cost, such as long processing times, high numbers of slack variables being required, and/or significantly increased complexity of solution. The present disclosure describes systems and methods useful in improving computational efficiency when solving
5 constrained quadratic problems. In particular, the presently described systems and methods may allow for specifying constraints without the requirement to convert them and include them in the objective function.

In order to allow some flexibility in solving while also ensuring that constraints are satisfied, a penalty value may be assigned to the constraints, allowing the weight that the
10 constraints are given to be varied, such that the constraint may be violated in some circumstances (such as at the start of a simulated annealing). When adding constraints to an objective function, a penalty value for the constraint may need to be selected without any guidance as to an appropriate penalty value in the given problem. This may result in the need to solve the problem multiple times with different penalty values to find better solutions. The
15 presently described systems and methods may also allow for selecting a penalty value without solving the problem multiple times with different penalty values, and for dynamic adjustment of the penalty value during the solving process.

The methods and systems described herein beneficially allow for computationally efficient solving optimization problems with constraints by handling the
20 constraints directly as opposed to converting them to part of an objective function. This may beneficially allow a processor to return feasible solutions to an optimization problem in a time efficient and scalable manner. This may also allow for the generation of multiple solutions in parallel by the processor. The contributions of the constraints may be handled implicitly through an update process that considers each constraint. Handling the constraint functions
25 directly and pulling energy values for the constraint functions through the computations may increase the efficiency of solutions. The methods and systems described herein may also advantageously allow for automatic and dynamic adjustment of penalty weights for the constraint functions, allowing for the search to be efficiently directed toward feasibility while returning better solutions. Many real-world problems, such as those solved in industry, are
30 problems having constraints on viable solutions. The methods and systems described herein may also allow for solution of problems having a set of variables with different types, such as a

mixture of binary and discrete variables. Sampling may be done based on a determination of the variable type and considering the current energy values and penalties. In some implementations, the methods and systems described herein may be used in combination with hybrid quantum computing techniques to provide improved solutions.

5 According to an aspect, there is provided a method of operation of a computing system to update a sample in an optimization algorithm to improve convergence to feasibility, the method being performed by a processor, the method comprising receiving a problem definition comprising a set of variables, an objective function defined over the set of variables, and one or more constraint functions, each of the constraint functions defined by at
10 least one variable of the set of variables, receiving sample values for the set of variables and a value for a progress parameter, for each variable of the set of variables determining a variable type for the variable, selecting a sampling distribution based on the variable type, determining an objective energy bias based on the sample value for the variable and one or more terms of the objective function that include the variable, determining one or more constraint energy
15 biases based on the sample value for the variable and each of the constraint functions defined by the variable, and sampling an updated value for the variable from the sampling distribution based on the objective energy bias, the one or more constraint energy biases, and the progress parameter, and returning an updated sample, the updated sample comprising the updated value for each variable of the set of variables.

20 According to other aspects the method may further comprise receiving a value for a penalty parameter, and sampling an updated value for the variable from the sampling distribution may further comprise sampling an updated value for the variable from the sampling distribution based on the value for a penalty parameter, receiving a value for the penalty parameter may comprise receiving a value for a Lagrange parameter that depends on
25 the value for the progress parameter, determining the variable type for the variable may comprise determining that the variable type is one of binary, discrete, integer, or continuous, determining that the variable type is one of binary, discrete, integer, or continuous may comprise determining that the variable type is binary, and wherein selecting a sampling distribution based on the variable type may comprise selecting a Bernoulli distribution,
30 determining that the variable type is one of binary, discrete, integer, or continuous may comprise determining that the variable type is discrete, and selecting a sampling distribution based on the variable type may comprise selecting a SoftMax distribution, determining that

the variable type is one of binary, discrete, integer, or continuous may comprise determining that the variable type is one of integer or continuous, and selecting a sampling distribution based on the variable type may comprise selecting a conditional probability distribution, wherein sampling an updated value from the sampling distribution may comprise slice
 5 sampling from the conditional probability distribution, and receiving the value for the progress parameter may comprise receiving an inverse temperature.

According to an aspect, there is provided a system for updating a sample in an optimization algorithm, the system comprising at least one non-transitory processor-readable medium that stores at least one of processor executable instructions and data and at least one
 10 processor communicatively coupled to the least one non-transitory processor-readable medium, which, in response to execution of the at least one of processor executable instructions and data performs a method as described herein.

According to an aspect, there is provided a method of operation of a computing system, the computing system comprising one or more processors, the method
 15 being performed by at least one of the one or more processors, the method comprising receiving a problem definition comprising a set of variables, an objective function defined over the set of variables, and one or more constraint functions, each of the constraint functions defined by at least one variable of the set of variables, initializing a sample solution to the objective function and a progress parameter, iteratively until a termination criteria is met
 20 incrementing a stage of an optimization algorithm, for each variable in the set of variables selecting an i^{th} variable from the set of variables, the i^{th} variable having a current value, calculating an objective energy bias for the objective function based on the current value of the i^{th} variable, calculating a constraint energy bias for each constraint function defined by the i^{th} variable based on the current value of the i^{th} variable, and sampling an updated value for
 25 the i^{th} variable based on the objective energy bias and the constraint energy bias, the updated value replacing the current value, incrementing a progress parameter, evaluating a termination criteria, and outputting a solution comprising the current values for the set of variables.

According to other aspects selecting an i^{th} variable from the set of variables
 30 may comprise selecting a binary variable, the binary variable having a current value and an alternative value, calculating an objective energy bias for the objective function based on the current value of the i^{th} variable may comprise calculating a difference in energy for the

objective function between the current value and the alternative value for the binary variable, calculating a constraint energy bias based on each constraint function defined by the i^{th} variable based on the current value of the i^{th} variable may comprise calculating a difference in energy for each constraint function defined by the binary variable between the current value
 5 of the binary variable and the alternative value for the binary variable, and sampling an updated value for the i^{th} variable based on the objective energy bias and the constraint energy bias may comprise sampling an updated value for the i^{th} variable based on the difference in energy values for the objective function and each constraint function defined by the i^{th} variable.

10 According to other aspects, the method may further comprise initializing a penalty parameter, adjusting the penalty parameter for each constraint function based on the difference in energy for the constraint function defined by the i^{th} variable and the progress parameter, and wherein calculating the difference in energy for each constraint function defined by the i^{th} variable includes penalizing each constraint function by the penalty
 15 parameter, incrementing a stage of an optimization algorithm for the objective function may include incrementing one of a simulated annealing or a parallel tempering algorithm, receiving a problem definition comprising an objective function and one or more constraint functions may comprise receiving a problem definition comprising a quadratic objective function and one or more quadratic equality or inequality constraint functions, receiving a problem
 20 definition comprising a set of variables may comprise receiving a problem definition comprising a set of one or more of binary, integer, or discrete variables, receiving a problem definition comprising a set of variables may comprise receiving a problem definition comprising one or more integer variables, and sampling an updated value for the i^{th} variable may comprise performing sampling from a conditional probability distribution, the i^{th} variable
 25 comprising an integer variable from the one or more integer variables, and the termination criteria may comprise one of a number of iterations, an amount of time, an average change in value limit, or a value of the progress parameter.

According to an aspect, there is provided a system for use in optimization, the system comprising at least one non-transitory processor-readable medium that stores at least
 30 one of processor executable instructions and data and at least one processor communicatively coupled to the least one non-transitory processor-readable medium, which, in response to

execution of the at least one of processor executable instructions and data, performs a method as described herein.

According to other aspects, the system may further comprise a quantum processor, and wherein, after performing the method, the at least one processor may instruct
5 the quantum processor to perform quantum annealing based on the outputted solution.

According to an aspect, there is provided a method of operation of a hybrid computing system, the hybrid computing system comprising a quantum processor and a classical processor, the method being performed by the classical processor, the method comprising receiving a constrained quadratic optimization problem, the constrained quadratic
10 optimization problem comprising a set of variables, an objective function defined over the set of variables, one or more constraint functions, each of the constraint functions defined by at least one variable of the set of variables, and a progress parameter for the optimization, the progress parameter comprising a set of values incrementing between an initial value and a final value, iteratively until the final value of the progress parameter is reached sampling a
15 sample set of values for the set of variables from an optimization algorithm, updating the sample set of values with an update algorithm comprising for each variable of the set of variables determining a variable type for the variable, selecting a sampling distribution based on the variable type, determining an objective energy bias based on a sample value for the variable from the sample set of values and one or more terms of the objective function that
20 include the variable, determining one or more constraint energy bias based on the sample value for the variable and each of the constraint functions defined by the variable, and sampling an updated value for the variable from the sampling distribution based on the objective energy bias, the one or more constraint energy biases, and the progress parameter, and returning an updated sample, the updated sample comprising the updated value for each
25 variable of the set of variables, incrementing the progress parameter, transmitting one or more final samples to a quantum processor, instructing the quantum processor to refine the samples, and outputting solutions.

According to other aspects, transmitting one or more final samples to a quantum processor may comprise transmitting pairs of samples to the quantum processor,
30 and wherein instructing the quantum processor to refine the samples comprises instructing the quantum processor to perform quantum annealing to select between the samples, and the

method may further comprise returning the outputted solutions as a sample set of values for the set of variables as an input to the optimization algorithm.

According to an aspect, there is provided a hybrid computing system, the hybrid computing system comprising a quantum processor and a classical processor, at least one non-transitory processor-readable medium that stores at least one of processor executable instructions and data, and at least one processor communicatively coupled to the least one non-transitory processor-readable medium, which, in response to execution of the at least one of processor executable instructions and data, performs a method as described herein.

According to an aspect, there is provided a method of operation of a computing system to direct a search space towards feasibility to improve performance of the computing system, the computing system comprising one or more processors, the method being performed by at least one of the one or more processors, the method comprising receiving a sample from an optimization, determining an energy value for one or more constraint functions, evaluating feasibility of the sample, if the sample is not feasible, increasing a penalty value, if the sample is feasible, decreasing a penalty value, and returning a penalty value to an optimization algorithm.

According to other aspects, the method may further comprise determining if a violation has been reduced in comparison with a previous sample and increasing an initial adjuster value if the violation has not been reduced and determining if a current best solution has been improved in comparison with a previous sample and increasing an initial adjuster value if the current best solution has not been improved.

According to an aspect, there is provided a method of operation of a computing system, the computing system comprising one or more processors, the method being performed by at least one of the one or more processors, the method comprising: receiving a problem definition comprising a set of variables, an objective function defined over the set of variables, and one or more constraint functions, each of the constraint functions defined by at least one variable of the set of variables, the set of variables comprising a first subset of variables having at least one variable that is one of binary, integer, discrete, or continuous and a second subset of variables having at least one variable that is continuous, initializing a sample solution to the objective function and a progress parameter, initializing a continuous problem defined over the second subset of variables, the continuous problem

comprising a linear programming model, iteratively until a termination criteria is met:
 incrementing a stage of an optimization algorithm, for each variable in the first subset of
 variables: determining a variable type for the variable, selecting a sampling distribution based
 on the variable type, determining an objective energy bias based on the sample value for the
 variable and one or more terms of the objective function that include the variable,
 5 determining one or more constraint energy biases based on the sample value for the variable
 and each of the constraint functions defined by the variable, sampling an updated value for
 the variable from the sampling distribution based on the objective energy bias, the one or
 more constraint energy biases, and the progress parameter, and returning an updated sample,
 10 the updated sample comprising the updated value for each variable of the set of variables,
 solving the linear programming model for the second subset of variables with the values for
 each variable in the first subset of variables fixed at the updated value, sampling an updated
 value for each variable in the second subset of variables based on the solved linear
 programming model, the updated value replacing a current value, updating the objective
 15 energy bias and the one or more constraint energy biases based on the updated values for the
 second subset of variables, incrementing the progress parameter, evaluating a termination
 criteria, and outputting a solution comprising the updated values for the set of variables.

According to other aspects, the method may further comprise initializing one
 or more penalty parameters, wherein sampling an updated value for the variable from the
 20 sampling distribution further comprises sampling an updated value for the variable from the
 sampling distribution based on at least one value for the one or more penalty parameters, and
 wherein sampling an updated value for each variable in the second subset of variables
 comprises sampling an updated value for each variable in the second subset of variables based
 on the solved linear programming model and at least one value for the one or more penalty
 25 parameters, initializing one or more penalty parameters may comprise initializing at least one
 Lagrange parameter that depends on the value for the progress parameter, determining the
 variable type for the variable may comprise determining that the variable type is one of
 binary, discrete, integer, or continuous, determining that the variable type is one of binary,
 discrete, integer, or continuous may comprise determining that the variable type is binary, and
 30 wherein selecting a sampling distribution based on the variable type comprises selecting a
 Bernoulli distribution, determining that the variable type is one of binary, discrete, integer, or
 continuous may comprise determining that the variable type is discrete, and wherein selecting

a sampling distribution based on the variable type comprises selecting a SoftMax distribution, determining that the variable type is one of binary, discrete, integer, or continuous may comprise determining that the variable type is one of integer and continuous, selecting a sampling distribution based on the variable type may comprise selecting a conditional probability distribution, sampling an updated value for the variable from the sampling distribution may comprise slice sampling from the conditional probability distribution, receiving the value for the progress parameter may comprise receiving an inverse temperature, incrementing a stage of an optimization algorithm for the objective function may include incrementing one of a simulated annealing or a parallel tempering algorithm, receiving a problem definition comprising an objective function and one or more constraint functions may comprise receiving a problem definition comprising a quadratic objective function and one or more quadratic equality or inequality constraint functions, and the termination criteria may comprise one of a number of iterations, an amount of time, an average change in value limit, or a value of the progress parameter.

According to an aspect, there is provided a system for use in optimization, the system comprising: at least one non-transitory processor-readable medium that stores at least one of processor executable instructions and data and at least one processor communicatively coupled to the least one non-transitory processor-readable medium, which, in response to execution of the at least one of processor executable instructions and data, performs a method as described herein.

According to other aspects, the system may further comprise a quantum processor, and, after performing a method as described herein, the at least one processor instructs the quantum processor to perform quantum annealing based on the outputted solution.

In other aspects, the features described above may be combined together in any reasonable combination as will be recognized by those skilled in the art.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWING(S)

In the drawings, identical reference numbers identify similar elements or acts. The sizes and relative positions of elements in the drawings are not necessarily drawn to scale. For example, the shapes of various elements and angles are not necessarily drawn to scale,

and some of these elements may be arbitrarily enlarged and positioned to improve drawing legibility. Further, the particular shapes of the elements as drawn, are not necessarily intended to convey any information regarding the actual shape of the particular elements and may have been solely selected for ease of recognition in the drawings.

5 Figure 1 is a schematic diagram of a hybrid computing system including a digital computer coupled to an analog computer, in accordance with the present systems, devices, and methods.

 Figure 2 is a schematic diagram of a portion of an exemplary superconducting quantum processor.

10 Figure 3 is a flow diagram of an example method for performing sampling in an optimization algorithm.

 Figure 4 is a flow diagram of an example method of operation of a computing system for finding a solution to a constrained problem.

15 Figure 5 is a flow diagram of another example method of operation of a computing system for finding a solution to a constrained problem.

 Figure 6 is a flow diagram of an example method of operation of a hybrid computing system.

 Figure 7 is an illustration of functions for constraints for equalities and inequalities.

20 Figure 8 is a flow diagram of an example method for adjusting penalty parameters to direct a search space toward feasibility.

 Figure 9 is a flow diagram of another example method for adjusting penalty parameters to direct a search space toward feasibility.

25 Figure 10 is a diagram of the domain of variable x_k used in calculating a partition function.

 Figure 11 is a diagram of the general case of a partition function for a segment s .

 Figure 12 is a diagram of an example probability distribution for a binary variable based on an energy bias and a progress parameter.

30 Figure 13 is a diagram of an example probability distribution function for continuous variables.

Figure 14 is a flow diagram of an example method of operation of a computing system for finding a solution to a constrained problem having continuous variables.

DETAILED DESCRIPTION

In the following description, certain specific details are set forth in order to
5 provide a thorough understanding of various disclosed implementations. However, one skilled
in the relevant art will recognize that implementations may be practiced without one or more
of these specific details, or with other methods, components, materials, etc. In other
instances, well-known structures associated with computer systems, server computers, and/or
communications networks have not been shown or described in detail to avoid unnecessarily
10 obscuring descriptions of the implementations.

Unless the context requires otherwise, throughout the specification and
claims that follow, the word “comprising” is synonymous with “including,” and is inclusive or
open-ended (*i.e.*, does not exclude additional, unrecited elements or method acts).

Reference throughout this specification to “one implementation” or “an
15 implementation” means that a particular feature, structure, or characteristic described in
connection with the implementation is included in at least one implementation. Thus, the
appearances of the phrases “in one implementation” or “in an implementation” in various
places throughout this specification are not necessarily all referring to the same
implementation. Furthermore, the particular features, structures, or characteristics may be
20 combined in any suitable manner in one or more implementations.

As used in this specification and the appended claims, the singular forms “a,”
“an,” and “the” include plural referents unless the context clearly dictates otherwise. It should
also be noted that the term “or” is generally employed in its sense including “and/or” unless
the context clearly dictates otherwise.

25 The headings and Abstract of the Disclosure provided herein are for
convenience only and do not interpret the scope or meaning of the implementations.

As an illustrative example, a superconducting quantum processor designed to perform adiabatic quantum computation and/or quantum annealing is used in the description that follows. However, as previously described, a person of skill in the art will appreciate that the present systems and methods may be applied to any form of quantum processor hardware (e.g., superconducting, photonic, ion-trap, quantum dot, topological, etc.) implementing any form of quantum algorithm(s) (e.g., adiabatic quantum computation, quantum annealing, gate/circuit-based quantum computing, etc.). A quantum processor may be used in combination with one or more classical or digital processors. Methods described herein may be performed by a classical or digital processor in communication with a quantum processor that implements a quantum algorithm.

Exemplary Computing System

Figure 1 illustrates a computing system 100 comprising a digital computer 102. The example digital computer 102 includes one or more digital processors 106 that may be used to perform classical digital processing tasks. Digital computer 102 may further include at least one system memory 122, and at least one system bus 120 that couples various system components, including system memory 122 to digital processor(s) 106. System memory 122 may store one or more sets of processor-executable instructions, which may be referred to as modules 124.

The digital processor(s) 106 may be any logic processing unit or circuitry (for example, integrated circuits), such as one or more central processing units ("CPUs"), graphics processing units ("GPUs"), digital signal processors ("DSPs"), application-specific integrated circuits ("ASICs"), programmable gate arrays ("FPGAs"), programmable logic controllers ("PLCs"), etc., and/or combinations of the same.

In some implementations, computing system 100 comprises an analog computer 104, which may include one or more quantum processors 126. Quantum processor 126 may include at least one superconducting integrated circuit. Digital computer 102 may communicate with analog computer 104 via, for instance, a controller 118. Certain

computations may be performed by analog computer 104 at the instruction of digital computer 102, as described in greater detail herein.

Digital computer 102 may include a user input/output subsystem 108. In some implementations, the user input/output subsystem includes one or more user input/output components such as a display 110, mouse 112, and/or keyboard 114.

System bus 120 may employ any known bus structures or architectures, including a memory bus with a memory controller, a peripheral bus, and a local bus. System memory 122 may include non-volatile memory, such as read-only memory ("ROM"), static random-access memory ("SRAM"), Flash NAND; and volatile memory such as random-access memory ("RAM") (not shown).

Digital computer 102 may also include other non-transitory computer- or processor-readable storage media or non-volatile memory 116. Non-volatile memory 116 may take a variety of forms, including: a hard disk drive for reading from and writing to a hard disk (for example, a magnetic disk), an optical disk drive for reading from and writing to removable optical disks, and/or a solid-state drive (SSD) for reading from and writing to solid state media (for example NAND-based Flash memory). Non-volatile memory 116 may communicate with digital processor(s) via system bus 120 and may include appropriate interfaces or controllers 118 coupled to system bus 120. Non-volatile memory 116 may serve as long-term storage for processor- or computer-readable instructions, data structures, or other data (sometimes called program modules or modules 124) for digital computer 102.

Although digital computer 102 has been described as employing hard disks, optical disks and/or solid-state storage media, those skilled in the relevant art will appreciate that other types of nontransitory and non-volatile computer-readable media may be employed. Those skilled in the relevant art will appreciate that some computer architectures employ nontransitory volatile memory and nontransitory non-volatile memory. For example, data in volatile memory may be cached to non-volatile memory. Or a solid-state disk that employs integrated circuits to provide non-volatile memory.

Various processor- or computer-readable and/or executable instructions, data structures, or other data may be stored in system memory 122. For example, system memory 122 may store instructions for communicating with remote clients and scheduling use of resources including resources on the digital computer 102 and analog computer 104. Also, for example, system memory 122 may store at least one of processor executable instructions or data that, when executed by at least one processor, causes the at least one processor to execute the various algorithms to execute instructions. In some implementations system memory 122 may store processor- or computer-readable calculation instructions and/or data to perform pre-processing, co-processing, and post-processing to analog computer 104.

System memory 122 may store a set of analog computer interface instructions to interact with analog computer 104. For example, the system memory 122 may store processor- or computer-readable instructions, data structures, or other data which, when executed by a processor or computer causes the processor(s) or computer(s) to execute one, more or all of the acts of the methods 300 (Figure 3) through 900 (Figure 9) and 1400 (Figure 14).

Analog computer 104 may include at least one analog processor such as quantum processor 126. Analog computer 104 may be provided in an isolated environment, for example, in an isolated environment that shields the internal elements of the quantum computer from heat, magnetic field, and other external noise. The isolated environment may include a refrigerator, for instance a dilution refrigerator, operable to cryogenically cool the analog processor, for example to temperature below approximately 1 K.

Analog computer 104 may include programmable elements such as qubits, couplers, and other devices (also referred to herein as controllable devices). Qubits may be read out via readout system 128. Readout results may be sent to other computer- or processor-readable instructions of digital computer 102. Qubits may be controlled via a qubit control system 130. Qubit control system 130 may include on-chip Digital to Analog Converters (DACs) and analog lines that are operable to apply a bias to a target device. Couplers that couple qubits may be controlled via a coupler control system 132. Coupler control system 132 may include tuning elements such as on-chip DACs and analog lines. Qubit control system 130 and coupler control system 132 may be used to implement a quantum

annealing schedule as described herein on analog processor 104. Programmable elements may be included in quantum processor 126 in the form of an integrated circuit. Qubits and couplers may be positioned in layers of the integrated circuit that comprise a first material. Other devices, such as readout control system 128, may be positioned in other layers of the integrated circuit that comprise a second material. In accordance with the present disclosure, a quantum processor, such as quantum processor 126, may be designed to perform quantum annealing and/or adiabatic quantum computation. Examples of quantum processors are described in U.S. Patent No. 7,533,068.

10 Exemplary Superconducting Quantum Processor

Figure 2 is a schematic diagram of a portion of an exemplary superconducting quantum processor 200, according to at least one implementation. Superconducting quantum processor 200 may be implemented within computing system 100 of Figure 1, forming all or part of quantum processor 126. The portion of superconducting quantum processor 200 shown in Figure 2 includes two superconducting qubits 201, and 202. Also shown is a tunable coupling (diagonal coupling) via coupler 210 between qubits 201 and 202 (i.e., providing 2-local interaction). While the portion of quantum processor 200 shown in Figure 2 includes only two qubits 201, 202 and one coupler 210, those of skill in the art will appreciate that quantum processor 200 may include any number of qubits and any number of couplers coupling information between them.

Quantum processor 200 includes a plurality of interfaces 221–225 that are used to configure and control the state of quantum processor 200. Each of interfaces 221–225 may be realized by a respective inductive coupling structure, as illustrated, as part of a programming subsystem and/or an evolution subsystem. Alternatively, or in addition, interfaces 221–225 may be realized by a galvanic coupling structure. In some implementations, one or more of interfaces 221–225 may be driven by one or more DACs. Such a programming subsystem and/or evolution subsystem may be separate from quantum processor 200, or may be included locally (i.e., on-chip with quantum processor 200).

In the operation of quantum processor 200, interfaces 221 and 224 may each be used to couple a flux signal into a respective compound Josephson junction 231 and 232 of qubits 201 and 202, thereby realizing a tunable tunneling term (the Δ_i term) in the system Hamiltonian. This coupling provides the off-diagonal σ^x terms of the Hamiltonian and these
 5 flux signals are examples of “delocalization signals”. Examples of Hamiltonians (and their terms) used in quantum computing are described in greater detail in, for example, U.S. Patent Application Publication No. 2014/0344322.

Similarly, interfaces 222 and 223 may each be used to apply a flux signal into a respective qubit loop of qubits 201 and 202, thereby realizing the h_i terms (dimensionless
 10 local fields for the qubits) in the system Hamiltonian. This coupling provides the diagonal σ^z terms in the system Hamiltonian. Furthermore, interface 225 may be used to couple a flux signal into coupler 210, thereby realizing the J_{ij} term(s) (dimensionless local fields for the couplers) in the system Hamiltonian. This coupling provides the diagonal $\sigma_i^z \sigma_j^z$ terms in the system Hamiltonian.

15 In Figure 2, the contribution of each of interfaces 221–225 to the system Hamiltonian is indicated in boxes 221a–225a, respectively. As shown, in the example of Figure 2, the boxes 221a–225a are elements of time-varying Hamiltonians for quantum annealing and/or adiabatic quantum computing.

While Figure 2 illustrates only two physical qubits 201, 202, one coupler 210,
 20 and two readout devices 251, 252, a quantum processor (e.g., processor 200) may employ any number of qubits, couplers, and/or readout devices, including a larger number (e.g., hundreds, thousands or more) of qubits, couplers and/or readout devices. Examples of superconducting qubits include superconducting flux qubits, superconducting charge qubits, and the like. In a superconducting flux qubit, the Josephson energy dominates or is equal to the charging
 25 energy. In a charge qubit this is reversed. Examples of flux qubits that may be used include radio frequency superconducting quantum interference devices, which include a superconducting loop interrupted by one Josephson junction, persistent current qubits, which include a superconducting loop interrupted by three Josephson junctions, and the like.

Constrained Quadratic Problems

Quadratic functions refer to polynomial functions with one or more variables that have at most second-degree interactions. Many real-world problems can be expressed as a combination of a quadratic function to be optimized and a number of constraints placed on the feasible outcomes for the variables. In other words, a quadratic function (also referred to herein as an objective function) defines interactions between the variables, subject to a constraint or set of constraints. In order to obtain optimal or near-optimal solutions to a given problem, the objective function can be minimized or maximized, subject to the constraints on feasible results. In the example implementations described below, the problems are structured such that minimization or lower energies correspond with improved solutions. However, it will be understood that minimization problems may alternatively be structured as maximization problems (for example by reversing the sign of the function), and that the principles below apply generally to situations where an objective function is extremized. While minimization is discussed for clarity below, it will be understood that similar principles could be applied to functions that are maximized, and the term “extremized” could be substituted for “minimized”, and the energy penalties described below would be applied to penalize directions away from improved solutions.

One method of addressing constraints when solving an optimization problem is to convert the constraints to part of the objective function using penalty terms, and in the case of inequalities, additional “slack” variables, and then attempt to minimize over the new objective function. However, in some implementations, this may result in a high computational cost, such as long processing times, high numbers of slack variables being required, and/or significantly increased complexity of solution. The present disclosure describes systems and methods useful in improving computational efficiency when solving constrained quadratic problems. In particular, the presently described systems and methods may allow for specifying constraints without the requirement to convert them and include them in the objective function.

In order to allow some flexibility in solving, which may beneficially increase the probability of finding a global optimum, while also ensuring that constraints are satisfied

by a final solution, a penalty value may be assigned to the constraints, allowing the weight that the constraints are given to be varied, such that the constraint may be violated in some circumstances (such as at the start of a simulated annealing). When adding constraints to an objective function, a penalty value for the constraint may need to be selected without any
 5 guidance as to an appropriate penalty value in the given problem. This may result in the need to solve the problem multiple times with different penalty values to find better solutions. The presently described systems and methods may also allow for selecting a penalty value without solving the problem multiple times with different penalty values, and for dynamic adjustment of the penalty value during the solving process.

10 The intended outcome of solving a constrained optimization problem is the optimization of some function, referred to herein as $f(x)$, subject to some inequality or equality constraints such as $g_c(x) \leq 0$. In a quadratic optimization problem, $f(x)$ and $g_c(x)$ are functions with at most quadratic polynomial interactions. The optimization can be expressed as $x = \arg \min_x f(x)$. This may be expressed in terms of

$$15 \quad f(x) = \sum_i h_i x_i + \sum_{i,j} J_{ij} x_i x_j$$

Subject to constraints

$$\sum A_{c,i} x_i + \sum_{i,j} B_{c,i,j} x_i x_j + b_c = 0$$

In binary implementations having linear equality constraints, the linear constraint can be expressed as a function:

$$20 \quad \delta_c = \sum_i A_{c,i} x_i + b_c = 0$$

$$\delta_{c,0} = \delta_c - x_i A_{c,i}$$

$$\delta_{c,1} = \delta_c + (1 - x_i) A_{c,i}$$

$$\Delta_E = \sum_c (|\delta_{c,1}|^n - |\delta_{c,0}|^n) \quad n = 0, 1, \text{ or } 2$$

In order to treat this constraint as a quadratic term, the left-hand side (the
 25 violation from the desired value) can be squared ($n = 2$). When the violations are small,

particularly where violations are less than one, it may be beneficial to instead use a linear term ($n = 1$) or a constant term ($n = 0$) instead. As discussed in further detail below, for each proposed update to variable x_i , the Δ_E for each constraint is determined. Increased energy values penalize the solutions that produce those increased values, as the desired outcome is a minimization of the energy of the system.

In implementations having linear inequality constraints, the functions are treated similarly. However, in this case, the violation will be penalized only where the function is positive, and not if the function is negative. So, for example, for an inequality expressed as $\sum_i A_{c,i} x_i + b_c \leq 0$, The violation will be penalized only where the error function is positive, otherwise no penalty will be applied.

In implementations having quadratic constraints, the functions may similarly be expressed by:

$$\begin{aligned}\delta_c &= \sum_i A_{c,i} x_i + \sum_{i,j} B_{c,i,j} x_i x_j + b_c = 0 \\ \delta_{c,0} &= \delta_c - x_i A_{c,i} - x_i \sum_j B_{c,i,j} x_j \\ \delta_{c,1} &= \delta_c + (1 - x_i) A_{c,i} + (1 - x_i) \sum_j B_{c,i,j} x_j \\ \Delta_E &= \sum_c (|\delta_{c,1}|^n - |\delta_{c,0}|^n) \quad n = 0, 1, \text{ or } 2\end{aligned}$$

As above, the value of n may be selected based on the value of the violation. Inequality constraints may be penalized only for a given range of values, as discussed in further detail below.

20

Figure 3 is a flow diagram of an example method 300 for performing sampling in an optimization algorithm as executed on a processor-based system. Method 300 may be executed on a hybrid computing system comprising at least one digital or classical processor and a quantum processor, for example hybrid computing system 100 of Figure 1, or by a classical computing system comprising at least one digital or classical processor. An example

25

implementation of method 300 is discussed in further detail below with reference to a constrained binary problem.

Method 300 comprises acts 302 to 322; however, a person skilled in the art will understand that the number of acts illustrated is an example, and, in some
 5 implementations, certain acts may be omitted, further acts may be added, and/or the order of the acts may be changed.

Method 300 starts at 302, for example in response to a call or invocation from another routine.

At 302, the processor receives a problem definition having a set of variables,
 10 an objective function defined over the set of variables, and one or more constraint functions, each of the constraint functions defined by at least one variable of the set of variables.

The problem definition may be expressed as:

$$\text{Minimize:} \quad \sum_i a_i x_i + \sum_{i \leq j} b_{ij} x_i x_j + c$$

$$\text{Subject to:} \quad \sum_i a_i^{(c)} x_i + \sum_{i \leq j} b_{ij}^{(c)} x_i x_j + c^{(c)} \leq 0, \quad c = 1, \dots, C_{ineq} \text{ and}$$

$$15 \quad \sum_i a_i^{(d)} x_i + \sum_{i \leq j} b_{ij}^{(d)} x_i x_j + c^{(d)} = 0, \quad d = 1, \dots, C_{eq},$$

Where $\{x_i\}_{i=1, \dots, N}$ represents the set of variables, the sets a_i , b_{ij} , and c are real values, and C_{ineq} and C_{eq} are the number of inequality and equality constraints, respectively.

At 304, the processor receives sample values for the set of variables and a
 20 value for a progress parameter. In some implementations, the progress parameter may be an inverse temperature. The sample values may be a predetermined starting value based on known properties of the problem, may be provided by another routine, may be randomly selected, or may be provided using other techniques as are known in the art.

At 306, the processor selects an i^{th} variable from the set of variables. This
 25 may, for example, be a first variable in the set of variables on a first iteration, a second variable in the set of variables on a second iteration, etc., until all of the variables within the

set of variables have been selected. In other implementations, the order of the variables may be randomized or selected according to other metrics as are known in the art.

At 308, the processor determines a variable type for the variable. The variable type may be, for example, one of binary, discrete, integer, and continuous. It will be
 5 understood that for a given problem definition, all of the variables in the set of variables may be a single type, or the problem may have a mixture of different types of variables.

At 310, the processor selects a sampling distribution based on the variable type. In some implementations, where the variable type is binary, the sampling distribution selected may be the Bernoulli distribution. In other implementations, where the variable type
 10 is discrete, the sampling distribution selected may be the SoftMax distribution. Alternatively, in some implementations, discrete variables may be included as binary variables using one-hot constraints and the sampling distribution may be the Bernoulli distribution. One-hot constraints refers to the transformation of categorical variables into binary variables by assigning a 0/1 or True/False dummy binary variable to each category. For example, discrete
 15 variables may be added as $\sum_{i \in D} x_i = 1$. In other implementations, where the variable type is integer, sampling may be performed by Gibbs sampling, that is, sampling on the conditional probability given the current state, as discussed in further detail below with reference to Figures 10 and 11. Alternatively, Gibbs sampling may be performed on all variable types, with the conditional probability function being determined by the variable type. It will be
 20 understood that in some implementations, integer variables may also be transformed to binary variables and sampling from the Bernoulli distribution may be performed. In some implementations, where the variable type is continuous, sampling on the conditional probability may be done by slice sampling, or the sampling distribution may be provided by a linear programming model, as discussed below in further detail.

At 312, the processor determines an objective energy bias acting on the variable under consideration given the current values of all of the other variables. As discussed above, an optimization problem may be structured with an objective function that defines an energy, and during optimization a processor returns solutions with the goal of reducing this energy. In some implementations, such as where the variable under
 30 consideration is a binary variable, an objective energy change (Δ_{E_l}) may be determined based on the energy of the previous solution and the energy of the current sample value, with a decrease in energy suggesting an improved solution.

At 314, the processor determines a constraint energy bias acting on the variable under consideration given the current values of all of the other variables and each of the constraint functions that involves the variable. A constraint energy bias may be defined by $\Delta_E = \sum_c (|\delta_{c,1}|^n - |\delta_{c,0}|^n)$ for binary variables, as discussed above. The penalty applied to
 5 each constraint may be determined by the magnitude of the violation.

In some implementations, the processor determines a total energy change based on the objective energy change and the constraint energy change. The constraint functions are indirectly included in this energy by the processor by the addition of an energy term considering the constraints, such as in the function $\Delta_{E_i, total} = \Delta_{E_i, objective} +$
 10 $\lambda (\delta_{c,1}^2 \Theta(\delta_{c,1}) - \delta_{c,0}^2 \Theta(\delta_{c,0}))$ for binary variables, as discussed above.

At 318, the processor samples an updated value for the variable from the sampling distribution based on the objective and constraint energy biases and the progress parameter. For example, in some implementations, the processor may sample from a distribution such that where the sample value was an improvement in energy and did not
 15 violate any constraint functions, the sample value is accepted, while if the sample value was not an improvement in energy and/or violated a constraint function, the sample value is only accepted with some probability that depends on the progress parameter, and otherwise a previous value is accepted. The progress parameter may allow for more violations at an early stage and fewer violations at a later stage. In other implementations, the sampled updated
 20 value may be sampled from a weighted distribution based on the total energy change and the progress parameter. For example, where the variable is a binary variable, the sampling distribution may be the Bernoulli distribution, and the variable may be sampled according to $x_i \sim B(1/(1 + e^{-\beta \Delta_{E_i}}))$, where Δ_{E_i} is the total energy change, and β is the progress parameter (e.g., the inverse temperature).

25 In other implementations, where Gibbs sampling is performed on binary variables, the conditional partition function is used. For binary variables, the conditional probability that a variable $x_k = 1$ is given by $p(x_k = 1) = Z_\beta(k)^{-1} e^{-\beta f(x_k=1|\{x_i\}_{i \neq k})}$, where β is the progress parameter. The conditional partition function for the objective function is therefore given by $Z_\beta(k) = e^{-\beta f(x_k=1|\{x_i\}_{i \neq k})} + e^{-\beta f(x_k=0|\{x_i\}_{i \neq k})}$. For the objective function,
 30 the objective energy bias is given by $\delta_{eff} = a_k + \sum_{i \neq k} b_{ik} x_i$, and the probability function is given by $p(x_k = 1) = \frac{1}{1 + e^{\beta \delta_{eff}}}$. An example probability distribution based on different

progress parameters is shown in Figure 12. For the adjacent constraints, the partition function can be written as:

$$Z_{\beta}(k) = \text{const} \left\{ \exp \left[-\beta(\delta_{eff} + \lambda_c(\delta_1^{(c)} - \delta_0^{(c)})) \right] + 1 \right\},$$

Where *const* is a constant prefactor and the probability is given by $p(x_k = 1) =$

$$\frac{1}{1 + e^{\beta(\delta_{eff} + \lambda_c(\delta_1^{(c)} - \delta_0^{(c)}))}}.$$

For discrete variables, Gibbs sampling may be performed based on the use of the one-hot constraints. Based on the expression $\text{const} + \sum_{d \in D_k} \delta_{eff}^d x_k^d$, where *d* is the case index, D_k is the set of cases for variable *k*, the partition function can be given as $Z_{\beta}(k) = \sum_{d \in D_k} e^{-\beta \delta_{eff}^d}$. Gibbs sampling may then be performed based on the values in different

segments.

For integer variables, the partition function can be given by $Z_{\beta}(k) =$

$(e^{-\beta \delta_{eff} \underline{x}_k} - e^{-\beta \delta_{eff} (\bar{x}_k + 1)}) (1 - e^{-\beta \delta_{eff}})^{-1}$, where $\underline{x}_k$ is the lower bound on variable x_k and $\bar{x}_k$ is the upper bound. As discussed in further detail below, when considering constraints for integer values, the binding values $\tilde{x}_k^{(c)}$, which are the values at which both sides of an

inequality are equal, are determined, and the partition function for each segment is given by $\sum_{n=\underline{x}_s}^{\bar{x}_s+1} \exp \left\{ -\beta \left[\delta_{eff} n + \sum_{c \in C} \lambda_c \delta_{eff}^{(c)} \left(n - \tilde{x}_k^{(c)} \right)^{\alpha} \right] \right\}$. λ_c is a penalty term for each constraint, used to penalize infeasible solutions. Penalty terms are discussed in further detail below. α is the penalization norm parameter, also discussed in further detail below. At 320, the processor evaluates if all of the variables of the set of variables have been considered. For example, where *n* variables are in the set of variables, and the processor starts with x_1 , then x_s , etc., the processor may monitor an incremental value between 1 and *n*, and when x_n is considered, the incremental value may be *n*, and the processor will evaluate that all of the variables of the set of variables have been considered. For continuous variables, sampling may be performed from a linear programming model or a slice sampling model, as discussed in further detail below. In some implementations, sampling may be performed first for other variable types as a first subset, and then performed for the continuous variables as a second subset. If all of the variables have not been considered, control passes back to act 306, and the next variable is selected. Once all of the variables have been considered, control passes to act 322.

At 322, the processor returns an updated sample with the updated value for each variable of the set of variables. Method 300 may then terminate until it is, for example, invoked again, or method 300 may repeat with new sample values from act 304. The sample output at 322 may be passed to another algorithm for further processing or may be returned
 5 as a solution to the problem. In some implementations, method 300 may be applied after a simulated annealing algorithm returns a solution, with method 300 being applied at 0 temperature, resulting in the system relaxing to the closest local minima.

In some implementations, method 300 may include receiving a value for a penalty parameter, and the total energy change may also depend on the value for the penalty
 10 parameter. Penalty parameters are discussed in further detail below. In some implementations, the penalty parameter may be a Lagrange parameter that depends on the value for the progress parameter.

An example pseudocode for method 300 in an example implementation with
 15 only binary variables is as follows:

```

Initialization;
 $\delta_c = \sum_i A_{c,i} x_i + \sum_{i,j} B_{c,i,j} x_i x_j$ ;
for i in variables do
   $\Delta_{E_i} = 0$ ;
  20 for c in neighborhood(i) do
     $\delta_{c,0} = \delta_c - x_i A_{c,i} - \sum_j x_i x_j B_{c,i,j}$ ;
     $\delta_{c,1} = \delta_c + A_{c,i}(1 - x_i) + (1 - x_i) \sum_j x_j B_{c,i,j}$ ;
     $\Delta_{E_i} = \Delta_{E_i} + \delta_{c,1}^2 \Theta(\delta_{c,1}) - \delta_{c,0}^2 \Theta(\delta_{c,0})$ ;
  end
  25  $x_i \sim B(1/(1 + e^{-\beta \Delta_{E_i}}))$ ;
end
```

As discussed above, the progress parameter may be an inverse temperature β . In some implementations, the inverse temperature may be incremented through a set
 30 range, such as starting at 0.1 and incrementing to 1. In other implementations, it may be beneficial to set the initial temperature (the highest temperature of the simulated annealing

algorithm and therefore the smallest inverse temperature), β_{min} , and the final temperature (the lowest temperature of the simulated annealing algorithm), β_{max} based on the probability of moving away from a local minimal. That is, the starting and ending temperatures should be selected to be sufficiently high that there is a significant probability of moving toward a less
5 desirable solution. In some implementations, the solutions returned by a processor may be sensitive to the initial temperature, and determining an initial temperature based on the specific problem may beneficially return improved solutions.

In some implementations, the energy bias contributed by the objective function and any constraints may be used to determine an initial inverse temperature
10 parameter that provides a significant probability (for example a 50% probability) of sampling a value that is not locally optimal. At a high temperature during the beginning of the algorithm, this probability may be selected to guarantee that even for variables having the strongest possible energy bias, the most unfavorable values still have a relatively large probability of being sampled. At a low temperature at the end of the algorithm, the probability may be
15 selected to guarantee that even for a variable with the smallest possible energy bias there is at least a relatively small probability of sampling the least favorable value. In other words, throughout the algorithm the probability of selecting unfavorable values starts at a predetermined relatively high probability, and then decreases, but is always non-zero.

Figure 4 is a flow diagram of an example method 400 of operation of a
20 computing system for finding a solution to an optimization problem or other problem that may be expressed as a constrained quadratic model, or for generating sample solutions that may be used as inputs to other algorithms. Method 400 may be executed on a hybrid computing system comprising at least one digital or classical processor and a quantum processor, for example hybrid computing system 100 of Figure 1, or by a classical computing
25 system comprising at least one digital or classical processor. Method 400 may be used in implementations with binary variables, or method 400 may include one or more additional acts to express other variable types as binary variables.

Method 400 comprises acts 402 to 430; however, a person skilled in the art will understand that the number of acts illustrated is an example, and, in some

implementations, certain acts may be omitted, further acts may be added, and/or the order of the acts may be changed.

Method 400 starts at 402, for example in response to a call or invocation from another routine or in response to an input by a user.

5 At 402, the processor receives a problem definition. The problem definition includes a set of variables, an objective function defined over the set of variables, and one or more constraint functions, each of the constraint functions defined by at least one variable of the set of variables. The problem definition may be received from a user (e.g., entered via an input device), transmitted from another processor, retrieved from memory, or provided as the
10 output of another process performed by the processor. The objective function may be a quadratic function, and the problem definition may define a quadratic optimization problem. The set of variables may contain one or more of continuous, discrete, binary, and integer variables. The constraint functions may include one or more quadratic equality or inequality constraint functions.

15 At 404, the processor initializes a sample solution to the objective function. The sample solution may be a random solution to the objective function. The random solution may be randomly selected from across the entire variable space, or the random solution may be selected within a limited range of the variable space based on known properties of the problem definition or other information. The sample solution may also be generated by
20 another algorithm or provided as input by a user.

 At 406, the processor initializes a progress parameter. The progress parameter may be a set of incrementally changing values that define an optimization algorithm. For example, the progress parameter may be an inverse temperature, and the inverse temperature may increment from an initial high temperature to a final low
25 temperature. In some implementations, an inverse temperature is provided as a progress parameter for a simulated annealing algorithm. The selection of an inverse temperature is described in further detail above.

At 408, the processor optionally initializes a penalty parameter. In some implementations the penalty parameter may be a Lagrange multiplier as discussed in further detail below. In other implementations the penalty parameter may be selected as a constant value or a value that depends on one or more other variables.

5 At 410, the processor optionally calculates a current value of each constraint function at the sample values. For example, constraint functions may be given as $\sum A_{c,i}x_i + b_c = 0$, or as inequalities, depending on the constraints provided.

 At 412, the processor increments a stage of the optimization algorithm, such as by providing the progress parameter to the optimization algorithm. Optimization
10 algorithms may include simulated annealing, parallel tempering, Markov Chain Monte Carlo techniques, branch and bound algorithms, and greedy algorithms, which may be performed by a classical computer. Optimization algorithms may also include algorithms performed by a quantum computer, such as quantum annealing, quantum approximate optimization algorithm (QAOA) or other noisy intermediate-scale quantum (NISQ) algorithms, quantum
15 implemented fault-tolerant optimization methods, or other quantum optimization algorithms. Quantum computers may include quantum annealing processors or gate model based processors. Examples of some implementations of optimization algorithms are described in U.S. provisional patent application number 62/951,749 and U.S. patent application publication number 2020/0234172. On successive iterations the incremented optimization algorithm may
20 provide samples.

 At 414, the processor selects an i^{th} variable from the set of variables. This may, for example, be a first variable in the set of variables on a first iteration, a second variable in the set of variables on a second iteration, etc., until all of the variables within the set of variables have been selected. The i^{th} variable has a current value from the sample
25 initialized at act 404. In some implementations the i^{th} variable may also have an alternative value. The alternative value may be provided at act 412 by the optimization algorithm or may be generated based on known properties of the i^{th} variable. For example, if the i^{th} variable is binary, and the current value for the i^{th} variable is 0, then the alternative value for the i^{th} variable will be 1.

At 416, the processor calculates an energy bias provided by the objective function based on the current value of the i^{th} variable. In implementations with alternative values, such as where the variable is binary, this may include calculating the difference in energy for the objective function between the current value of the i^{th} variable and the
 5 alternative value for the i^{th} variable. As discussed above, an objective energy change (Δ_{E_i}) for binary variables may be determined based on the objective function evaluated for the i^{th} variable taking the current value and the objective function evaluated for the i^{th} variable taking the alternative value.

At 418, the processor calculates an energy bias provided by the constraints
 10 that involve the i^{th} variable. There may be one or more constraint functions that contribute to the constraint energy bias. In implementations with alternative values, this may include calculating the difference in energy for each constraint function defined by the i^{th} variable between the current value of the i^{th} variable and the alternative value for the i^{th} variable. Calculating the difference in energy for each constraint function defined by the i^{th} variable may
 15 include penalizing each constraint function by its respective penalty parameter. For example, in the case of a binary variable with a linear equality constraint, the difference in energy for the constraint function may be given as $\delta_{c,0} = \delta_c - x_i A_{c,i}$ and $\delta_{c,1} = \delta_c + (1 - x_i) A_{c,i}$, with a difference in energy given by $\Delta_E = \sum_c (|\delta_{c,1}|^n - |\delta_{c,0}|^n)$, as discussed above.

At 420, the processor samples a value for the i^{th} variable based on the
 20 objective and constraint energy biases. In implementations with alternative values, this may include sampling based on the difference in energy values for the objective function and each constraint function defined by the i^{th} variable, as calculated at acts 416 and 418. As discussed above, the sampled value may be sampled from a weighted distribution based on the total energy change and the progress parameter, such as the Bernoulli distribution for a binary
 25 variable, the SoftMax distribution for a discrete variable, the conditional probability distribution for an integer variable, or a linear programming or slice sampling model for a continuous variable.

At 422, the processor evaluates if all of the variables of the set of variables have been considered. If all of the variables have not been considered, control passes back to

act 414, and the next variable is selected. Once all of the variables have been considered, control passes to act 424. In some implementations, the processor may incrementally consider each variable in turn of the set of variables and evaluate that all of the variables of the set of variables have been considered when the last variable of the set of variables has
5 been considered.

At 424, the processor increments the progress parameter. For example, if the progress parameter is an inverse temperature, and was initialized at t_0 at act 406, the progress parameter may be incremented to t_1 during a first iteration at 424. In some implementations, t_1 may be a temperature that is lower than t_0 . On successive iterations, the temperature may
10 be further decreased until a final temperature is reached.

At 426, optionally, the penalty parameter for each constraint may be adjusted. In some implementations the penalty parameter may be adjusted based on the change in energy for the constraint function defined by the i^{th} variable and the progress parameter. In other implementations, the penalty parameter may be adjusted as described in methods 800
15 and 900, discussed in further detail below.

At 428, the processor evaluates one or more termination criteria. In some implementations, the termination criteria may be a value of the progress parameter. In other implementations, the termination criteria may include a number of iterations, an amount of time, a threshold average change in energy between updates, a measure of the quality of the
20 current values for the variables, or other metrics as are known in the art. If the termination criteria are not met, the method continues with act 412. Acts 412-428 are repeated iteratively until the termination criteria is met. In some implementations, the set of variables sampled at act 420 may be received as input by the optimization algorithm at act 412, and the samples may be modified by the optimization algorithm before passing to act 414. As discussed above,
25 incrementing a stage of an optimization algorithm for the objective function may include incrementing a simulated annealing or parallel tempering algorithm. In other implementations the optimization algorithm may be a MCMC algorithm or a greedy algorithm. Other optimization algorithms include branch and bound algorithms, quantum annealing, quantum approximate optimization algorithm (QAOA) or other noisy intermediate-scale

quantum (NISQ) algorithms, quantum implemented fault-tolerant optimization methods, or other quantum optimization algorithms.

If the one or more termination criteria has been met, control passes to 430, where a solution is output. At 430, method 400 terminates, until it is, for example, invoked
5 again. The solution output at act 430 may be passed to other algorithms, such as a quantum annealing algorithm.

In some implementations, after outputting a solution at act 430, method 400 may begin again with a new sample being initialized at act 404. In some implementations, method 400 may be performed multiple times in parallel starting from different initialized
10 samples or a series of randomly generated samples. In some implementations, the solutions may be paired, and binary problems may be built to evaluate the series of solutions using Cross-Boltzmann updates as described in U.S. provisional patent application number 62/951,749.

The methods described herein use optimization algorithms, and may, in some
15 implementations, use simulated annealing to provide candidate solutions. Simulated annealing probabilistically decides to accept a candidate solution based both on the change in energy of the candidate solution and the stage in the simulated annealing. At early stages of the simulated annealing, candidate solutions that do not provide lower energies are more likely to be accepted, while at later stages of the simulated annealing, candidate solutions that
20 do not provide lower energies are more likely to be rejected. This may be thought of as evolving a system from a high temperature to a low temperature. The probability of accepting a candidate solution will depend on a probability acceptance function that depends both on the energies of the current solution and the candidate solution and a time varying parameter, often represented or referred to as temperature. Another alternative is parallel tempering
25 (also referred to as replica exchange Markov chain Monte Carlo). Similar to simulated annealing, parallel tempering begins at random initial solutions, and exchanges candidate solutions based on temperature and energies. Other optimization algorithms include MCMC algorithms, greedy algorithms, branch and bound algorithms, quantum annealing, quantum approximate optimization algorithm (QAOA) or other noisy intermediate-scale quantum

(NISQ) algorithms, quantum implemented fault-tolerant optimization methods, or other quantum optimization algorithms.

The methods described herein allow a solver to implicitly consider constraint functions and penalize violations without the requirement to convert constraint functions to
5 part of the objective function and determine appropriate penalty values. The method considers the interaction of each variable with all connected variables, and in addition, the method considers the interaction of each variable with each constraint. The method may also adjust the strength of penalties within the function until each constraint is satisfied. For each
10 variable, the methods may perform an update specific to the variable type to allow different types of variables to be included.

Figure 5 is a flow diagram of an example method 500 of operation of a computing system for finding a solution to an optimization problem or other problem that may be expressed as a constrained quadratic model and having binary variables. It will be
15 understood that similar principles may be applied to other types of variables discussed herein, either by converting the other variable types to binary variables, or by addressing the energy biases and sampling distributions as discussed in other methods herein. Method 500 may be executed on a hybrid computing system comprising at least one digital or classical processor and a quantum processor, for example hybrid computing system 100 of Figure 1, or by a
20 classical computing system comprising at least one digital or classical processor.

Method 500 comprises acts 502 to 524; however, a person skilled in the art will understand that the number of acts illustrated is an example, and, in some implementations, certain acts may be omitted, further acts may be added, and/or the order of the acts may be changed.

25 Method 500 starts at 502, for example in response to a call or invocation from another routine.

At 502, the processor receives a problem definition. The problem definition includes a set of variables, an objective function defined over the set of variables, and one or

more constraint functions, each of the constraint functions defined by at least one variable of the set of variables. The problem definition may be received from a user (e.g., via an input device), transmitted from another processor, retrieved from memory, or provided as the output of another process performed by the processor. The objective function may be a
5 quadratic function, and the problem definition may define a quadratic optimization problem. The set of variables may contain one or more of continuous, discrete, binary, and integer variables.

At 504, the processor generates two candidate values for each variable of the objective function. Example methods for generating candidate values are described in U.S.
10 provisional patent application number 62/951,749. The candidate values may be generated by an optimization algorithm, may be randomly generated, or a random solution may be selected, and alternative values based on the random solution may be generated. A solution may be randomly selected from across the entire variable space or may be selected within a range of the variable space based on known properties of the problem definition or other
15 information. For example, if the i^{th} variable is binary, and the current value for the i^{th} variable is 0, then the alternative value for the i^{th} variable will be 1. For non-binary variables, an alternative value may be generated by sampling from a known distribution. In some implementations the processor may use a native sampler, such as a Gibbs sampler, in the space of integer or continuous variables to select the alternative values. It will be understood
20 that alternative values may be generated by a variety of algorithms and may depend on the variable type and known parameters of the problem.

At 506, the processor selects an i^{th} variable from the set of variables. This may, for example, be a first variable in the set of variables on a first iteration, a second variable in the set of variables on a second iteration, etc., until all of the variables within the
25 set of variables have been selected.

At 508, the processor calculates the difference in energy for the objective function based on the two candidate values for the i^{th} variable generated at 504. An objective energy change (ΔE_i) may be determined based on the objective function evaluated for the i^{th} variable at each of the candidate values.

At 510, the processor calculates the difference in energy for each constraint function between the two candidate values for the i^{th} variable. In some implementations, the difference in energy may be given by $\Delta_E = \sum_c (|\delta_{c,1}|^n - |\delta_{c,0}|^n)$, as discussed above.

At 512, the processor samples a value for the i^{th} variable based on the
 5 difference in energy values for the objective function and each constraint function defined by the i^{th} variable. As discussed above, the sampled value may be sampled from a weighted distribution based on the total energy change and the progress parameter, such as the Bernoulli distribution for a binary variable.

At 514, the processor evaluates if all of the variables of the set of variables
 10 have been considered. In some implementations, the processor may incrementally consider each variable of the set of variables in turn and evaluate that all of the variables of the set of variables have been considered when the last variable of the set of variables has been considered. If all of the variables have not been considered, control passes back to act 506, and the next variable is selected. Once all of the variables have been considered, control
 15 passes to act 516.

At 516, the processor stores energy values for each constraint function based on the sampled values.

At 518, the processor optionally adjusts a penalty value applied to the energy calculation for the change in energy values for the constraint function. As discussed in further
 20 detail below, where a constraint is not satisfied, the penalty value may be increased, while where a constraint is satisfied, the penalty value may be decreased. Implementations of automatic adjustment of a penalty value is discussed below with respect to methods 800 and 900. In other implementations, the penalty value may be adjusted by a fixed value in response to a constraint being satisfied or violated.

At 520, the processor evaluates one or more termination criteria. Termination
 25 criteria may include a value for a progress parameter, a number of iterations, an amount of time, a threshold average change in energy between updates, a measure of the quality of the current values for the variables, or other metrics as are known in the art. If the termination

criteria is not met, the method continues with act 522. Acts 504-522 are repeated iteratively until the termination criteria is met. If the termination criteria has been met, control passes to 524, where a solution is output. At 524, method 500 terminates, until it is, for example, invoked again.

5 At 522, the processor increments an optimization algorithm for the objective function to provide a new set of accepted values for the set of variables. The optimization algorithm may start from the sampled values from act 512 and generate an alternative solution in order to provide two candidate values for each variable. Optimization algorithms may include simulated annealing, parallel tempering, Markov Chain Monte Carlo techniques, 10 branch and bound algorithms, and greedy algorithms, which may be performed by a classical computer. Optimization algorithms may also include algorithms performed by a quantum computer, such as quantum annealing, quantum approximate optimization algorithm (QAOA) or other noisy intermediate-scale quantum (NISQ) algorithms, quantum implemented fault-tolerant optimization methods, or other quantum optimization algorithms. Quantum 15 computers may include quantum annealing processors or gate model based processors. Examples optimization algorithms are described in U.S. provisional patent application number 62/951,749 and U.S. patent application publication number 2020/0234172.

Figure 6 is a flow diagram of an example method 600 of operation of a hybrid 20 computing system for finding a solution to an optimization problem or other problem that may be expressed as a constrained quadratic model. Method 600 may be executed on a hybrid computing system comprising at least one digital or classical processor and a quantum processor, for example hybrid computing system 100 of Figure 1.

Method 600 comprises acts 602 to 608; however, a person skilled in the art 25 will understand that the number of acts illustrated is an example, and, in some implementations, certain acts may be omitted, further acts may be added, and/or the order of the acts may be changed.

Method 600 starts at 602, for example in response to a call from another routine.

At 602, the processor receives a constrained quadratic model (CQM) problem (also referred to herein as a constrained quadratic optimization problem), the constrained
5 quadratic optimization problem having a set of variables, an objective function defined over the set of variables, one or more constraint functions, each of the constraint functions defined by at least one variable of the set of variables, and a progress parameter for the optimization, the progress parameter comprising a set of values incrementing between an initial value and a final value.

10 At 604, the processor generates sample solutions for the CQM problem using optimization algorithm 604a and sampling algorithm 604b. These methods may be similar to methods 300, 400, and 500 described herein. In some implementations, act 604 may include iteratively, until the final value of a progress parameter is reached: sampling a sample set of values for the set of variables from an optimization algorithm, updating the sample set of
15 values with an update algorithm by, for each variable of the set of variables: determining an objective energy change based on the sample value for the variable and one or more terms of the objective function that include the variable, determining a constraint energy change based on the sample value for the variable and each of the constraint functions defined by the variable, determining a total energy change based on the objective energy change and the
20 constraint energy change, determining a variable type for the variable, selecting a sampling distribution based on the variable type, and sampling an updated value for the variable from the sampling distribution based on the total energy change and the progress parameter, and then returning an updated sample, the updated sample comprising the updated value for each variable of the set of variables and incrementing the progress parameter. In other
25 implementations, act 604 may include any other methods as described herein, for example, method 1400 of Figure 14.

At 606, the one or more final samples generated after reaching the final value of the progress parameter, or another termination criteria as discussed above, may be transmitted to a quantum processor, and the processor may instruct the quantum processor

to refine the samples. In some implementations, transmitting the one or more final samples to a quantum processor includes transmitting pairs of samples to the quantum processor (at 606a), and instructing the quantum processor to refine the samples by performing quantum annealing to select between the samples (at 606b and 606c).

5 At 608, the processor outputs solutions to the CQM problem. In some implementations, these solutions may be returned to a user (e.g., via an output device). In other implementations, the solutions may be passed to another algorithm for refinement, or may be returned to act 604 as a sample set of values as an input to the optimization algorithm. At 608 method 600 terminates, unless it is iterated, or until it is, for example, invoked again.

10 As discussed above with respect to method 600, in some implementations, the obtained samples may be further refined by a quantum computer, for example as part of a hybrid algorithm that uses cluster contractions. Cluster contraction hybrid algorithms are disclosed in more details in US Patent Application Publication No. 2020/0234172. In some implementations, the solutions may be paired, and binary problems may be built to evaluate
15 the series of solutions using Cross-Boltzmann updates as described in U.S. provisional patent application number 62/951,749. The initial solution may be improved by using a hybrid computing system such as hybrid computing system 100 of Figure 1, comprising at least one classical or digital computer and a quantum computer.

 In a Cross-Boltzmann update, two possible solutions may be found or provided
20 with two different values for a given variable. A new binary variable is defined that selects between the two different values. This can then be reduced to a Quadratic Unconstrained Binary Optimization (QUBO) problem that may be optimized. In some implementations, the resulting QUBO problem may be sent to a quantum processor, and solutions may be generated using quantum annealing. A classical or digital computer may use Cross-Boltzmann
25 updates to build a binary problem that may then be solved by the quantum computer. Cross-Boltzmann updates use the energy associated with updating two variables, both individually, and dependent on one another, to provide a Hamiltonian that expresses the decision whether to update those two variables. The digital processor may select two candidate values and may then send this update to the quantum processor as an optimization problem.

A QUBO problem is defined using an upper-diagonal matrix Q , which is an $N \times N$ upper triangular matrix of real weights, and x , a vector of binary variables, to minimize the function:

$$f(x) = \sum_i Q_{i,i} x_i + \sum_{i < j} Q_{i,j} x_i x_j$$

5 where the diagonal terms $Q_{i,i}$ are the linear coefficients and the nonzero off-diagonal terms are the quadratic coefficients $Q_{i,j}$.

This can be expressed more concisely as $\min_{x \in \{0,1\}^n} x^T Q x$.

In scalar notation, the objective function expressed as a QUBO is as follows:

$$E_{qubo}(a_i, b_{i,j}; q_i) = \sum_i a_i q_i + \sum_{i < j} b_{i,j} q_i q_j$$

10

Inequalities

As discussed above, the contribution of the constraints to the change in energy can be represented as $\Delta_{E_i} = \Delta_{E_i} + \delta_{c,1}^2 \Theta(\delta_{c,1}) - \delta_{c,0}^2 \Theta(\delta_{c,0})$. $\Theta(\delta)$ is a function that addresses inequality constraints. Where an inequality constraint is not violated, $\Theta(\delta)$ will be zero, such that that term does not contribute to the energy. Referring to Figure 7, $\Theta(\delta)$ is selected such that for equalities, the energy penalty increases away from zero, as shown in graph 702. For inequalities, $\Theta(\delta)$ is zero as long as the inequality is not violated, as shown in graph 704.

15

For example, for an inequality expressed as:

20

$$\sum_i A_{c,i} x_i + b_c \leq 0$$

Penalizing the violation according to $\delta_c \Theta(\delta_c)$, where $\Theta(\delta)$ is the function shown in graph 704 will only penalize the violation if the value of the constraint is positive.

Lagrange Parameters

As discussed above, the change in energy contributed by both the objective function and the constraints can be represented as $\Delta E_i = \Delta E_i + \lambda \delta_{c,1}^2 \Theta(\delta_{c,1}) - \lambda \delta_{c,0}^2 \Theta(\delta_{c,0})$, where λ is a Lagrange parameter (also referred to herein as a Lagrange multiplier). The Lagrange parameter acts as a penalty parameter, with the value of the penalties assigned to each function being adjusted iteratively throughout the solution process in order to direct the search spaces toward feasibility. When a variable update returns a non-viable solution as described above, the penalty value may be increased. When a variable update returns a viable solution, the penalty may be relaxed to increase the likelihood that an optimal solution can be found.

The Lagrange multiplier may be adjusted during the optimization by evaluating the violation of the constraint functions. If a given constraint is violated, the associated Lagrange multiplier is increased. If a given constraint is satisfied, then the associated Lagrange multiplier is decreased.

As in the implementations discussed above, a CQM problem can be expressed as:

$$\begin{aligned} \text{minimize } f &= \sum_i a_i x_i + \sum_{i,j} b_{i,j} x_i x_j \\ g_m &= \sum_i a'_{i,m} + \sum_{i,j} b'_{i,j,m} + c_m \leq 0 \quad \forall m \in M \end{aligned}$$

where x_i are the problem variables (which may be binary, discrete, integer, continuous, etc.) and M is set of constraints. a_i is the linear bias for variable i and $b_{i,j}$ are quadratic biases between variables i and j in the objective function. Similarly, $a'_{i,m}$ $b'_{i,j,m}$ are linear and quadratic biases for constraint m .

In simulated annealing, for each variable, the magnitude of the change in the objective value is evaluated for where the simulated state of the variable changes from its current value.

$$\delta E^k = \delta_f^k + \sum_m \mu_m^k (\delta v_m^k)^2$$

where δ_f^k is the change in the objective value and δv_m^k is the change due to violation in constraints in iteration k of the simulated annealing algorithm. μ_m^k is the Lagrange multiplier for each constraint. It may be beneficial to adjust μ_m^k to achieve a balance between the weight given to the objective function and the weight given to the constraints in order to attempt to minimize the objective function while also satisfying the constraints. Selecting overly high values for μ_m^k may satisfy constraints prematurely and result in a less optimal solution to the objective function, while overly low values may result in solutions being produced that violate one or more constraints, also known as infeasible solutions.

In one implementation for updating μ_m^k , four metrics may be used to determine the update: i) the violation on each constraint, ii) the distance from binding for each constraint if the constraint is not violated, iii) the total violations across all of the constraints, and iv) the total distance from binding across all of the constraints. This may be represented as:

$$\mu_m^k = \mu_m^{k-1} \left(1 + \frac{\alpha_m^k v_m^k}{\left| \sum_{m'} v_m^{k^2} \right|} + \frac{\gamma_m^k s_m^k}{\left| \sum_{m'} s_{m'}^k \right|} \right)$$

where v_m^k and s_m^k are the violation and the distance from binding for constraint m respectively, and are calculated using:

$$v_m^k = \begin{cases} 0 & \text{if } g_m(x^k) \leq 0 \\ g_m(x^k) & \text{if } g_m(x^k) > 0 \end{cases} \quad \begin{matrix} \text{if constraint is lower equality} \\ \text{if constraint is lower equality} \\ |g_m(x^k)| & \text{if constraint is equality} \end{matrix}$$

$$s_m^k = \begin{cases} g_m^k(x) & \text{if } g_m(x^k) \leq 0 \\ 0 & \text{if } g_m(x^k) \geq 0 \end{cases} \quad \begin{matrix} \text{if constraint is lower equality} \\ \text{if constraint is lower equality} \\ 0 & \text{if constraint is equality} \end{matrix}$$

α_m^k and γ_m^k are two multipliers that are adjusted during the optimization using a global damping factor (df) and a local increment method.

Global damping is used to stabilize the penalty values after finding the first feasible solution. The parameters are damped using a damp factor df :

$$df = 1 / \sqrt{k - r}$$

$$\alpha_m^k = \alpha_m^{k-1} df$$

$$\gamma_m^k = \gamma_m^{k-1} df$$

where α^0 and γ^0 are initial adjuster values, k is the iteration number, and r is the iteration number for the first feasible solution.

5 A local increment method is used such that when the problem becomes infeasible and fails to improve the total violation, α_0 is increased exponentially by a constant factor (e.g., 1.2) and when the problem becomes feasible and fails to improve the current best possible solution, γ_0 is increased exponentially by a constant factor (e.g., 1.2) after t iterations defined as:

$$10 \quad t = \sqrt{\log_{scale} (\beta_{max} / \beta_{min})}$$

where β_{min} , β_{max} and $scale$ are simulated annealing parameters.

Figures 8 and 9 are flow diagrams of example methods 800 and 900 for adjusting penalty parameters to direct the search space toward feasibility, which may advantageously improve performance of a processor-based system executing the methods 15 800 and 900. Methods 800 and 900 may be implemented as subprocess of simulated annealing, or as part of an optimization algorithm. As discussed below, method 900 starts with initialization of the Lagrange initial parameters and multipliers, and in particular the initial Lagrange values μ , α and γ . In addition, tin and tf counters are initialized to count the number of iterations that the search fails to make progress in lowering infeasibility (total violations), and best feasible solution, respectively. tin is the number of iterations that the search stays in the infeasible region without lowering infeasibility (lowering total violation), while tf is the number of iterations that the search stays in feasible region without improving the best 20 feasible solution. t , as discussed above, is the number of iterations that the algorithm waits before increasing α and γ .

25 During the update method of simulated annealing, for a given state, its current objective, and constraint left hand side energy, the system checks the feasibility of the solution, distance of constraints from binding, and their violations. Then the system adjusts

the search toward feasibility or infeasibility according to improvement in best possible solution and total violations. Finally, Lagrange multipliers are adjusted based on the new value for α , γ , violation/constraints bindings and Lagrange multipliers from previous iteration.

In some implementations, the Lagrange parameters may also be varied
5 directly with the progression of the simulated annealing along the progress parameter, such as inverse temperature. At the start of the simulated anneal, the Lagrange parameters may be smaller, and the penalty assigned to violations may be smaller. Near the end of the simulated anneal, the penalty assigned to violations may increase such that the constraints may be satisfied in the final solutions.

10 Figure 8 is a flow diagram of an example method 800 for directing a search space towards feasibility in an optimization algorithm, which may advantageously improve the performance of a processor-based system in performing optimization. Method 800 may be executed on a hybrid computing system comprising at least one digital or classical processor and a quantum processor, for example hybrid computing system 100 of Figure 1, or by a
15 classical computing system comprising at least one digital or classical processor.

Method 800 comprises acts 802 to 822; however, a person skilled in the art will understand that the number of acts illustrated is an example, and, in some implementations, certain acts may be omitted, further acts may be added, and/or the order of the acts may be changed.

20 Method 800 starts at 802, for example in response to a call or invocation from another routine.

At 802, the processor receives a sample from an optimization algorithm, such as a simulated annealing algorithm as discussed above. In some implementations, method 800 may be called by any one of methods 300, 400, 500, and 600 as discussed above, or
25 method 1400 discussed below. Method 800 may be used to adjust penalty parameters, such as at act 426 of method 400, or act 518 of method 500.

At 804, the processor determines an energy value or bias for one or more constraint functions. In some implementations, the processor may evaluate the constraint

functions for the given sample. In other implementations, such as in the methods discussed above, an energy value for one or more constraint functions may be calculated by another function and passed to act 804, such as by act 418 of method 400 or act 516 of method 500.

At 806, the processor evaluates the feasibility of the sample. This may, for example, be determined by the value received for the energy value for the one or more constraint functions. In some implementations, a function may be provided, as discussed above with reference to Figure 7, that zeros out energy values for feasible samples. Where a constraint is violated, the sample is considered to be infeasible.

At 808, control passes to 810 where the sample is not feasible, and control passes to 812 where the sample is feasible.

At 810, the processor increases the penalty value for a constraint that is violated for a non-feasible sample. In some implementations, the penalty value may be incremented by a constant value at each iteration where a non-feasible sample is returned. As only feasible samples are desired, incrementing the penalty value until the constraint functions are sufficiently weighted that the algorithm returns only feasible solutions is beneficial.

At 812, the processor decreases the penalty value for a feasible sample. As having too much weight on a constraint may result in feasible solutions that are less than optimal, the penalty value may be decreased where feasible samples are returned. As discussed below with respect to method 900, a damping factor may be used to converge the penalty value to a low value that returns feasible solutions. In some implementations, the penalty value may be decreased by a constant value at each iteration where a feasible sample is returned until a non-feasible sample is returned. The penalty value may then be increased until a feasible sample is returned, and then maintained at that penalty value for remaining iterations.

At 814, the processor optionally determines if a violation has been reduced in comparison with a previous sample. For example, this may be done by comparing constraint violation functions as discussed above. If the total energy penalty contributed by the

constraint functions increases, the violation has been increased. If the total energy penalty contributed by the constraint functions decreases, the violation has been reduced. In some implementations, the violation may also remain the same, and as such has not been reduced. If the violation has been reduced, control passes to 818, otherwise to 822.

5 At 818, the processor increases an initial adjuster value. For example, as discussed above, α_0 may be increased exponentially by a constant factor (e.g., 1.2).

 At 816, the processor optionally determines if a current best solution has been improved in comparison with a previous sample. If the current best solution has been improved, control passes to 822, otherwise to 820. For example, this may be done by
10 comparing the energy of the objective function in the previous iteration to the energy of the objective function

 At 820, the processor increases an initial adjuster value. For example, as discussed above, γ_0 may be increased exponentially by a constant factor (e.g., 1.2).

 At 822, the processor outputs a penalty value to be returned to the
15 optimization algorithm. Method 800 may then terminate until it is, for example, invoked again.

 Figure 9 is a flow diagram of an example implementation of method 800 for directing a search space towards feasibility in an optimization algorithm, in the form of method 900, which may advantageously improve the performance of a processor-based
20 system in performing optimization. Figure 9 demonstrates how Lagrange multipliers may be automatically adjusted during a simulated annealing procedure for solving a CQM problem. Figure 9 illustrates the main simulated annealing procedure 902, along with acts 904-922 which illustrate a procedure for automatic adjustment of multipliers.

 At 904, the system initializes the Lagrange parameters. In the example
25 implementation of Figure 9, the Lagrange parameter μ_m^0 is initialized for iteration $k=0$ with initial adjusters α^0 and γ^0 set to 2 and 0.05 respectively. t represents a number of iterations that initial adjusters α^0 and γ^0 remain at the same value when the problem is infeasible and fails to improve the total violation, or when the problem is feasible and fails to improve the

best possible solution. $\beta_{\min}$, $\beta_{\max}$ and $scale$ are simulated annealing parameters provided by the simulated annealing algorithm. tin and tf are counters used to count the number of iterations in the infeasible and feasible regime during which the sampling has failed to make progress and are initially set to zero.

5 At 906, the system receives an energy for a constraint function and a current state of a variable from an optimization or update algorithm. For example, a sample value generated by one of methods 300, 400, 500, 600, or 1400 may be passed to the system, along with an energy for a constraint function found by one of those methods. In some implementations, the system may receive an energy for a constraint function and a current
10 state of a variable from a simulated annealing algorithm.

 At 908, the system evaluates a damp factor, violations, and bindings. A damp factor may be provided as an initial parameter or may be adjusted in response to fluctuations between feasible and infeasible values in previous iterations. Violations may be evaluated from constraint functions, such as discussed above. For example, a constraint may be
15 evaluated and multiplied by a function as discussed with respect to Figure 7. The binding may be calculated as a distance from binding for each constraint where the constraint is not violated, and/or the total distance from binding across all constraints.

 At 910, the system determines if the current sample is feasible, that is, if any of the constraints are violated. This may be determined based on the energy of the constraint,
20 as discussed above. If one or more of the constraints are violated, the method passes to act 912, while if the sample is feasible (i.e., no constraints are violated), the method passes to 914.

 At 912, the system determines if the violation has been reduced. As above, this may be determined based on the energy of the constraint. If the violation has been
25 reduced, control passes to 916, with no change to the counter values (tin and tf remain at zero). If the violation has not been reduced, control passes to 918.

 At 918, counter tin is incremented by one, and compared to t . If tin is greater than or equal to t , the value of initial adjuster α^0 is increased (for example by some

constant factor, such as an exponent of 1.2). If the value of initial adjuster α^0 is increased, tf is reset to zero. Control is then passed to 916.

At 914, the system determines if the best solution has been improved, that is, if the energy of the objective function has been decreased, or a more optimal solution has been found. If the solution has been improved, control passes to 916, with no change to the counter values (tf and tf remain at zero). If the best solution has not been improved, control passes to 920.

At 920, counter tf is incremented by one, and compared to t . If tf is greater than or equal to t , the value of initial adjuster γ^0 is increased (for example by some constant factor, such as an exponent of 1.2). If the value of initial adjuster γ^0 is increased, tf is reset to zero. Control is then passed to 916.

At 916, the system determines the effective values for α_m^k and γ_m^k based on a dampening factor df as described above. Once a feasible solution is found, the change in parameters is damped using the damp factor, which may, for example, depend on the current iteration number and the iteration number where the first feasible solution was found ($df = 1 / \sqrt{k - r}$). α_m^k and γ_m^k may be evaluated according to $\alpha_m^k = \alpha_m^{k-1} df$ and $\gamma_m^k = \gamma_m^{k-1} df$.

At 922, the system generates an updated Lagrange multiplier for each constraint based on if the constraint is satisfied, and the effective values for α_m^k and γ_m^k by evaluating the equation for μ_m^k discussed above: $\mu_m^k = \mu_m^{k-1} (1 + \frac{\alpha_m^k v_m^k}{|\sum_{m'} v_m^k|} + \frac{\gamma_m^k s_m^k}{|\sum_{m'} s_m^k|})$.

The updated Lagrange multipliers are then passed back to 902, and may be used in other algorithms, such as being passed back to a stage of a simulated annealing algorithm.

Example with Pseudocode

In some implementations, the problem to be solved may involve only binary variables. One example implementation of such a binary problem having binary variables x_1, x_2 will now be discussed. It will be understood that the example problem is representative of one possible

example implementation only, and that other implementations may go beyond what is discussed below.

The processor receives a constrained quadratic model (CQM) problem having an objective function such as $f(x_1, x_2)$, and constraint functions such as $g(x_1, x_2)$ and $h(x_1)$. The
5 constraint functions may include equalities or inequalities.

The processor receives or initializes a sample. The sample may be generated by another algorithm, may be received as an input, may be sampled from a distribution, or may be randomly generated. For the purposes of this example, a random sample is initialized as $\bar{x}_{\text{random}} = [0, 1]$. In some implementations, a pair of samples may be received.

10 The processor receives or initializes a progress parameter. In some implementations, the progress parameter may be an inverse temperature. The processor may, for example, initialize an inverse temperature as t_0 . For example, in some implementations, simulated annealing may be used to perform optimization. The temperature of the system may be set to some high temperature initially and may be incremented to some lower temperature. This
15 may have the effect of allowing the acceptance of higher energy solutions at the beginning of the optimization, reducing the probability of remaining in a local minimum, and preventing the acceptance of higher energy solutions near the end of the optimization.

The processor may optionally receive or initialize penalty parameters. For example, the processor may initialize Lagrange parameters as discussed in further detail above as a
20 function of the progress parameter and the magnitude of the energy violation, e.g., $\lambda_c(t, \delta_c)$.

The processor computes the current value of each constraint based on the sample that was received or initialized. In the given example of $\bar{x}_{\text{random}} = [0, 1]$, the processor computes $g(0,1)$, $h(0)$.

The processor then considers each variable in turn. The variable has a current value
25 provided by the initialized sample, and the processor may either determine an alternative value based on properties of or restrictions on the variable or may receive an alternative value as input or from another algorithm. For example, given that x_1 is a binary variable, and the sampled value from $\bar{x}_{\text{random}}$ is 0, the alternative value for x_1 must be 1. It will be understood that in the case of variable types other than binary, alternative values may be received with
30 the initial sample, or may be generated by another algorithm. The processor considers the

objective function and determines the energy terms contributed by the objective function in the neighborhood of the variable. The processor also considers the constraint functions that involve the variable for the current value of the variable and the alternative value of the variable. In the given example, the processor may begin with x_1 and find the value of the linear terms of the objective function $f(x_1, x_2)$ that involve x_1 for the initialized sample value (0) and an alternative value (1). This may include pairwise interaction terms between x_1 and x_2 and the coefficients of terms that involve only x_1 . It will be understood that in larger examples there may be a variety of pairwise interaction terms between different variables. An energy value, such as Δ_E may then be set to the energy contributed by these terms for x_1 .

For each constraint that involves x_1 , an energy value for the constraint is determined by the processor based on the initially calculated constraint values $g(0,1)$, $h(0)$, and the pairwise terms for $g(1,1)$, the alternative value for x_1 that involve x_1 . In the given example, the constraint energy may be calculated by the processor as $\delta_{x1,1} = g(1,1) + h(1)$ and $\delta_{x1,0} = g(0,1) + h(0)$. These terms are then added to the current energy value as $\Delta_E = \Delta_{E,x1} + \lambda_{t0,\delta x1,1} \delta_{x1,1}^2 \Theta(\delta_{x1,1}) - \lambda_{t0,\delta x1,0} \delta_{x1,0}^2 \Theta(\delta_{x1,0})$, where $\Theta(\delta)$ is a function for inequality constraints, as discussed in further detail above with reference to Figure 7. Where an inequality constraint is not violated, $\Theta(\delta)$ will be set to zero, such that that term does not contribute to the energy.

A sample value for variable x_1 is then generated by the processor based on Δ_E and t_0 . The sampling may vary depending on the type of variable, and may, for example, sample from a distribution, or generate samples in another way. In some implementations, if the sampled value does not violate any constraints, the sampled value is accepted, while if the sampled value does violate constraints, the sampled value is accepted with a probability that depends on t_0 . This value then becomes $x_{1,t0}$, for example, $x_{1,t0} = 1$.

The processor then considers x_2 and repeats the same to sample a value for $x_{2,t0}$.

Once all of the variables have been considered and a new sample value has been generated, the processor then increments the progress parameter. The generated sample $\bar{x}_{t0} = [x_{1,t0}, x_{2,t0}]$ may be passed back to take the place of $\bar{x}_{random}$, or some other processing operation may be performed. For example, in the case of simulated annealing, the generated sample $\bar{x}_{t0}$ may be passed back to a simulated annealing algorithm, which may sample new values starting from $\bar{x}_{t0}$, with the new values being returned to the presently discussed algorithm in the place of $\bar{x}_{random}$. The processor increments over the progress parameter,

generating an updated sample for each value of the progress parameter, such as from an initial inverse temperature to a final inverse temperature. The penalty parameter may also be updated at each increment of the progress parameter.

At the final inverse temperature, the generated sample $\bar{x}_{\text{final}} = [x_{1,\text{tf}}, x_{2,\text{tf}}]$ is returned.

- 5 Sampling from an initial sample over a progress parameter may be repeated multiple times, starting from a new randomly generated $\bar{x}_{\text{random}}$ for each repeat in the example above from the initial to final inverse temperature. This may be done in parallel for multiple starting samples. This results in a set of samples provided by a set of $\bar{x}_{\text{final}}$ values.

An example pseudocode for the above discussed simple binary problem is

10 provided below:

Receive objective function $f(x_1, x_2)$, constraints $g(x_1, x_2)$, $h(x_1)$
 Initialize a sample $\bar{x}_{\text{random}} = [0, 1]$
 Initialize penalty parameters $\lambda_c(t, \delta_c)$
 Initialize $t = t_0$

15

Compute current value of each constraint based on $\bar{x}_{\text{random}}$ to give $g(0,1)$, $h(0)$

For each t :

20

For each variable:

For x_1 :

Determine alternative value. $x_1 = 0$, could alternatively be 1 (binary).

$\Delta_{E,x1}$ = linear bias = the terms of $f(x_1, x_2)$ that involve x_1

$\Delta_{E,x1}$ = terms of $f(1,1)$ - terms of $f(0,1)$ that involve x_1

25

For each constraint c in neighborhood(x_1):

$\delta_{x1,1} = g(1,1) + h(1)$ – compute only pairwise terms that involve x_1 in $g(x_1, x_2)$ for $g(1,1)$.

$\delta_{x1,0} = g(0,1) + h(0) - g(0,1)$ and $h(0)$ are already known from above.

30

$\Delta_E = \Delta_{E,x1} + \lambda_{t(0,\delta_{x1,1})} \delta_{x1,1}^2 \Theta(\delta_{x1,1}) - \lambda_{t(0,\delta_{x1,0})} \delta_{x1,0}^2 \Theta(\delta_{x1,0})$

Sample a value for variable x_1 using Δ_E and t_0 :

Example: Returns $x_{1,t0} = 1$

35

Update $\delta_c = g(1,1) + h(1)$

For x_2 :

Alternative value: $x_2 = 0$.

$\Delta_{E,x2}$ = terms of $f(0,1)$ - terms of $f(0,0)$ that involve x_2

40

For each constraint c in neighborhood(x_1):

$\delta_{x2,0} = g(0,0)$ – compute only pairwise terms that involve x_2 in $g(x_1, x_2)$
 for $g(0,0)$.
 $\delta_{x2,1} = g(0,1) - g(0,0)$ is already known from above.
 $\Delta_E = \Delta_{E,x2} + \lambda_{t0,\delta_{x2,0}} \delta_{x2,0}^2 \Theta(\delta_{x2,0}) - \lambda_{t0,\delta_{x2,1}} \delta_{x2,1}^2 \Theta(\delta_{x2,1})$
 5
 Sample a value for variable x_2 using Δ_E and t_0 :
 Example: Returns $x_{2,t0} = 1$
 Update $\delta_c = g(1,1)$
 10
 Increment inverse temperature – Set $t = t_1$.
 Either pass to optimization algorithm and receive samples back or repeat for t_1
 directly.
 15

The methods and systems described herein beneficially allow for
 computationally efficient solving optimization problems with constraints by handling the
 constraints directly as opposed to converting them to part of an objective function. This may
 beneficially allow a processor to return feasible solutions to an optimization problem in a time
 20 efficient and scalable manner. This may also allow for the generation of multiple solutions in
 parallel by the processor. The contributions of the constraints may be handled implicitly
 through an update process that considers each constraint. Handling the constraint functions
 directly and pulling energy values for the constraint functions through the computations may
 increase the efficiency of solutions. The methods and systems described herein may also
 25 advantageously allow for automatic and dynamic adjustment of penalty weights for the
 constraint functions, allowing for the search to be efficiently directed toward feasibility while
 returning better solutions. Many real-world problems, such as those solved in industry, are
 problems having constraints on viable solutions. The methods and systems described herein
 may also allow for solution of problems having a set of variables with different types, such as a
 30 mixture of binary and discrete variables. Sampling may be done based on a determination of
 the variable type and considering the current energy values and penalties. In some
 implementations, the methods and systems described herein may be used in combination
 with hybrid quantum computing techniques to provide improved solutions.

35 Integer Variables

As discussed above, constrained problems may be defined over different variable types, such as binary, discrete, integer, and continuous. The variable types may determine the sampling distribution selected by the processor for sampling updated values for those variables. In some implementations, where the variable type is binary, the sampling
5 distribution selected may be the Bernoulli distribution. In other implementations, where the variable type is discrete, the sampling distribution selected may be the SoftMax distribution.

Where the variable type is integer (or continuous), there may not be a single well-defined distribution that may be selected. Many optimization problems include integer variables, and constraints on the values those integer variables may take. In order to support
10 integer variables in a constrained problem solver as discussed above, it may be beneficial to provide “native” support for integer variables. An update for integer variables considering their constraints and Lagrangian parameters is discussed below.

One technique for implementing integer updates includes temporarily relaxing the integer variable to a continuous variable, computing the gradient due to the objective and
15 Lagrangian relaxation of the constraints that involve that variable, proposing an update due to that gradient, and probabilistically accepting or rejecting it, similar to the process described above. However, these types of updates for integer variables may result in high algorithmic complexity (on the order of the number of adjacent constraints) where there is complexity in the constraints and/or many overlapping constraints. Alternatively, an integer variable,
20 defined with a bounded range of size R , could be implemented similarly to a binary variable by transforming the integer variable into approximately $\log(R)$ binary variables. This could then be updated by selecting the Bernoulli distribution as the sampling distribution and proceeding with updates as described above for binary variables. However, this type of update requires a selection of the best way to transform an integer variable into a binary variable. Some
25 examples include the base 2 representation or a Gray code. With $\log(R)$ binary variables, sampling may take exponential time, resulting in a run time on the order of R .

It may be beneficial to perform integer updates by examining the adjacent constraints for a given integer variable and dividing the range of the integer variable into a series of segments, with each segment having the same form of constraint penalty. A segment
30 may then be randomly chosen based on the integral of the objective and constraint penalties in that segment. A new value can be sampled for the integer from within the segment. While

the method below is discussed in the context of integer variables, it may also be extended for use with continuous variables as discussed below.

This process will take $O(M \log(M) + Mt)$ time, where M is the number of constraints and t is the runtime needed to compute the partition function or sample on a segment. In the case where the integer variable only has a linear term (or terms) in the objective function, $t = O(1)$, the update described as follows would be expected to take $O(M \log(M))$ time. If the integer variable has a quadratic term in the objective, t is expected to be $O(\log(r))$, where r is the range of the segment, and in some cases may be $O(1)$. This is beneficially expected to be faster than the other methods discussed above, such as relaxing integer variables to continuous variables or transforming into integer variables.

When sampling values for an integer variable, Gibbs sampling may be performed, that is, sampling on the conditional probability given the current state. This is described in further detail below. As discussed above, in the context of solving a CQM problem, both the objective function and the constraint functions are considered when sampling new values. Therefore, to update a given integer variable, sampling may be performed on the conditional probability given the current state or the presently accepted values of the variables. The conditional probability of the objective function may be computed analytically as described below.

It will be understood that while quadratic functions include squared variables, some of the equations in the implementation described herein may not work for a squared integer variable. In those implementations, a new variable may be defined. For example, where an integer variable x_1 is squared in the objective function or in a constraint, a new variable x_2 may be defined, and the square may be replaced with $x_1 x_2$, with the additional constraint $x_2 = x_1$. It will be understood that this substitution may be required in the equations used above, such as where variables are described with indexes i and j , and for the case $i=j$. It will be understood that other substitutions may similarly be made to address squared variables. In other implementations, the equations below may be modified. An alternative implementation to address squared (quadratic) variables without the addition of extra variables is also discussed below using slice sampling.

The constraints are considered by computing the binding values. Assuming that the effective linear bias on the variable for a constraint adjacent to the variable is greater than zero, there are three possible outcomes: the binding value may be less than the lower

bound on the variable, the binding value may be between the upper bound and the lower bound for the variable, and the binding value may be greater than the upper bound on the variable. In the case where the binding value is greater than the upper bound on the variable, the constraint is satisfied for all possible values of the variable and the partition function is
 5 given by the conditional partition function that includes only the conditional objective function. In the case where the binding value is less than the lower bound on the variable, all possible values of the variable violate the constraint and must be penalized. Where the effective linear bias on the variable is less than zero, these situations are reversed. As discussed above, a penalty may be applied as a constant or may be varied according to the
 10 magnitude of the violation and the stage of the optimization algorithm.

In the case where the binding value is between the upper and lower bounds on the variable, the domain may be divided into two segments as shown in Figure 10. As above, Figure 10 assumes that the effective linear bias is greater than zero. In the case where the effective linear bias is less than zero, the result will be inverted. The first segment is
 15 defined by the lower bound of the variable and the integer that is less than or equal to the binding value, and the second segment is defined by the integer that is equal to or greater than the binding value and the upper bound. The partition function may then be calculated as the sum of the partition functions of these two segments. The first segment can use the same partition function as the case where the binding value is greater than the upper bound on the
 20 variable, that is, the conditional partition function for the objective function. The second segment can use the same partition function as the case where the binding value is less than the lower bound on the variable, with the penalization applied. In the case where the effective linear bias on the variable is less than zero, the situation will be the same as above, but with the cases swapped. This may then be used to generalize to multiple constraints.

25 Given the functions for the objective functions and the constraints, a new value may be sampled by Gibbs sampling based on the conditional probability given the current state.

As discussed above, a constrained quadratic model (CQM) is defined by an objective function and a set of constraints. The CQM has a set of variables x_i , which may, for
 30 example, be combinations of binary, integer, discrete, and continuous variables, or other types of variables. In some implementations, one or more of the variables are integer variables.

In an implementation, an objective function is defined, for example, $f(x) = \sum_i a_i x_i + \sum_{i < j} w_{i,j} x_i x_j$. A set of M constraints of the form $g_c(x) = \sum_i a_{i,c} x_i + \sum_{i < j} w_{i,j,c} x_i x_j + C_c \leq, \geq, \text{ or } = 0$ is provided. A constraint $g_c(x)$ is considered to be “adjacent” for a given variable x_i if either $a_{i,c}$ or $w_{i,j,c}$ for any j is non-zero. The integer

5 variables in the CQM may be given with lower and/or upper bounds, which may be defined as $lb[k]$ (lower bound) and $ub[k]$ (upper bound) for a variable x_k . The bounds may alternatively be represented as constraints.

To update a given integer variable x_k , Gibbs sampling is performed on the conditional probability given the current state of all other variables. Considering the objective

10 function, when updating integer variable x_k , the conditional partition function is given by:

$$Z_{\beta}^{(k)} = \sum_{n=lb[k]}^{ub[k]} \exp[-\beta f(n|\{x_i\}_{i \neq k})] \quad (1)$$

and the conditional objective function can be written as:

$$15 \quad \exp[-\beta f(n|\{x_i\}_{i \neq k})] = \exp[-\beta A] \exp[-\beta \delta_{eff} n],$$

where the constant A includes all the terms in $f(x)$ that do not involve the variable x_k , $A = f(x) - \delta_{eff} x_k$, and δ_{eff} is the effective linear bias of the variable x_k , $\delta_{eff} = a_k + \sum_{i \neq k} w_{ik} x_i$.

When inserted in the partition function, the first term of the right-hand side of

20 the equation ($\exp[-\beta A]$) becomes an irrelevant pre-factor and the partition function can be taken as:

$$Z_{\beta}^{(k)} = \sum_{n=lb[k]}^{ub[k]} \exp[-\beta \delta_{eff} n].$$

In the case where $\delta_{eff} > 0$, the partition function can be computed analytically using:

$$25 \quad \sum_{n=0}^{\infty} x^n = \frac{1}{1-x}, |x| < 1,$$

And yields:

$$Z_{\beta}^{(k)} = (e^{-\beta \delta_{eff} lb[k]} - e^{-\beta \delta_{eff} (ub[k]+1)}) (1 - e^{-\beta \delta_{eff}})^{-1}. \quad (2)$$

In the case where $\delta_{eff} < 0$, the partition function can be analytically calculated using the same series identity as above with the substitution $m = -n$, to give:

$$Z_{\beta}^{(k)} = (e^{-\beta |\delta_{eff}| ub[k]} - e^{-\beta |\delta_{eff}| (lb[k]-1)}) (1 - e^{-\beta |\delta_{eff}|})^{-1}. \quad (3)$$

Considering the constraints, all constraints are expressed as inequality constraints, with equality constraints being implemented as a combination of two inequalities, and any constraint c adjacent to a variable x_k can be written as:

$$g_c(x_k | \{x_i\}_{i \neq k}) = A^{(c)} + \delta_{eff}^{(c)} x_k \leq 0, \quad (4)$$

and the binding values $\tilde{x}_k^{(c)}$ can be computed such that

$$g_c(x_k | \{x_i\}_{i \neq k}) = A^{(c)} + \delta_{eff}^{(c)} \tilde{x}_k^{(c)} = 0.$$

Where $\delta_{eff}^{(c)} > 0$, all of the values $n > \tilde{x}_k^{(c)}$ will need to be suppressed, while if $\delta_{eff}^{(c)} < 0$, all of the values $n < \tilde{x}_k^{(c)}$ need to be suppressed. Assuming that $\delta_{eff}^{(c)} > 0$, there are

three different possibilities:

1. $\tilde{x}_k^{(c)} < lb[k]$,
2. $lb[k] \leq \tilde{x}_k^{(c)} \leq ub[k]$,
3. $\tilde{x}_k^{(c)} > ub[k]$

Considering case 3, the constraint is satisfied for all possible values of x_k and no values need to be suppressed. The conditional partition function is given by equation (1).

Considering case 1, all possible values of the variable x_k violate the constraint and should therefore be penalized. The partition function can be written as:

$$Z_{\beta}^k = \sum_{n=lb[k]}^{ub[k]} \exp \left\{ -\beta [\delta_{eff} n + \lambda_c \delta_{eff}^{(c)} (n - \tilde{x}_k^{(c)})^{\alpha}] \right\},$$

Where $\alpha \geq 0$ allows the use of different penalties for violating the constraint.

In some implementations, $\alpha = 0, 1$, or 2 may be used, corresponding respectively to L_0 , L_1 , and L_2 penalties as discussed above. Note that the L_α norm of the violation in the equation should be $(\delta_{eff}^{(c)}(n - \tilde{x}_k^{(c)}))^\alpha$, which would mean that L_0 would have all constraints with the

5 same importance independent of the relative magnitude of violation of each constraint. The effective bias term $\delta_{eff}^{(c)}$ may therefore be excluded from the exponent as above.

Considering case 2, the domain of variable x_k can be divided in two segments as shown in Figure 10: $lb[k] \leq x_k \leq \lfloor \tilde{x}_k^{(c)} \rfloor$ and $\lceil \tilde{x}_k^{(c)} \rceil \leq x_k \leq ub[k]$. In Figure 10, diagonal line 1002 originating from $\tilde{x}_k^{(c)}$ denotes the direction and slope (given by $\delta_{eff}^{(c)}$) of the

10 penalization.

The partition function can be calculated as the sum of the partition functions in the two segments. For the first segment, the result in case 3 can be used, and in the second segment, the result of case 1 can be used, with both adapted to the segment domains. The partition function is then:

$$15 \quad Z_\beta^k = \sum_{n=lb[k]}^{\lfloor \tilde{x}_k^{(c)} \rfloor} \exp[-\beta \delta_{eff} n] + \sum_{n=\lceil \tilde{x}_k^{(c)} \rceil}^{ub[k]} \exp\{-\beta[\delta_{eff} n + \lambda_c \delta_{eff}^{(c)} (n - \tilde{x}_k^{(c)})^\alpha]\}.$$

Considering the case with $\delta_{eff}^{(c)} < 0$, cases 1 and 3 will be swapped, as well as the segments in case 2.

Generalizing the partition function to the case of multiple constraints, the ordered set of binding values $\{\tilde{x}_k^{(c)}\}_{c=1,\dots,C}$ defines a set of segments $\{s \equiv [x_s, x_{s+1}]\}$. As

20 above, the partition function can be computed over a segment and the partition functions of all segments can be summed. For each segment the contributing constraints are tracked. These are given by the constraints c for which $\delta_{eff}^{(c)} > 0$ and $\tilde{x}_k^{(c)} \leq x_s$ and also those for which $\delta_{eff}^{(c)} < 0$ and $\tilde{x}_k^{(c)} \geq x_{s+1}$.

Denoting this set of constraints as C_s , the partition function for a segment s is

25 given by:

$$Z_{\beta}^{k,(s)} = \sum_{n=x_s}^{x_{s+1}} \exp \left\{ -\beta \left[\delta_{eff} n + \sum_{c \in C_s} \lambda_c \delta_{eff}^{(c)} \left(n - \tilde{x}_k^{(c)} \right) \right] \right\}. \quad (4)$$

This general case is illustrated in Figure 11.

For the L_0 penalization, where $\alpha = 0$, the partition function becomes:

$$\begin{aligned} 5 \quad Z_{\beta}^{k,(s)} &= \exp \left\{ -\beta \sum_{c \in C_s} \lambda_c \delta_{eff}^{(c)} \right\} \sum_{n=x_s}^{x_{s+1}} \exp \{ -\beta \delta_{eff} n \} \\ &= \\ &\begin{cases} \exp \{ -\beta \sum_{c \in C_s} \lambda_c \delta_{eff}^{(c)} \} (e^{-\beta \delta_{eff} x_s} - e^{-\beta \delta_{eff} (x_{s+1}+1)}) (1 - e^{-\beta \delta_{eff}})^{-1}, & \text{if } \delta_{eff} > 0, \\ \exp \{ -\beta \sum_{c \in C_s} \lambda_c \delta_{eff}^{(c)} \} (e^{-\beta |\delta_{eff}| x_{s+1}} - e^{-\beta |\delta_{eff}| (x_s-1)}) (1 - e^{-\beta |\delta_{eff}|})^{-1}, & \text{if } \delta_{eff} < 0. \end{cases} \end{aligned} \quad (5)$$

For the L_1 penalization, an analytic expression of the partition function can be provided, as the linear dependence of the violation can be incorporated in the objective function bias. This obtains:

$$Z_{\beta}^{k,(s)} = \exp \left\{ \beta \sum_{c \in C_s} \lambda_c \delta_{eff}^{(c)} \tilde{x}_k^{(c)} \right\} \sum_{n=x_s}^{x_{s+1}} \exp \left\{ -\beta \left[\delta_{eff} + \sum_{c \in C_s} \lambda_c \delta_{eff}^{(c)} \right] n \right\}.$$

The results for the L_0 penalization can be used by changing the pre-factors and setting $\delta_{eff} \rightarrow \delta_{eff} + \sum_{c \in C_s} \lambda_c \delta_{eff}^{(c)}$.

15 The partition function for L_0 and L_1 can therefore be written in a uniform way by defining, for each segment s , a linear penalty LP_s with offset ω and slope σ as:

$$LP_s[\omega] = \begin{cases} \sum_{c \in C_s} \lambda_c \delta_{eff}^{(c)} & \text{if } \alpha = 0 \\ -\sum_{c \in C_s} \lambda_c \delta_{eff}^{(c)} \tilde{x}_k^{(c)} & \text{if } \alpha = 1 \end{cases} \quad (6)$$

And

$$20 \quad LP_s[\sigma] = \begin{cases} \delta_{eff} & \text{if } \alpha = 0 \\ \left[\delta_{eff} + \sum_{c \in C_s} \lambda_c \delta_{eff}^{(c)} \right] & \text{if } \alpha = 1 \end{cases} \quad (7)$$

And write the partition function as:

$$\begin{aligned}
 Z_{\beta}^{k,(s)} &= e^{-\beta\omega} \sum_{n=x_s}^{x_{s+1}} e^{-\beta\sigma n} \\
 &= \begin{cases} e^{-\beta\omega} (e^{-\beta\sigma x_s} - e^{-\beta\sigma(x_{s+1}+1)}) (1 - e^{-\beta\sigma})^{-1} & \text{if } \sigma > 0 \\ e^{-\beta\omega} (e^{-\beta|\sigma|x_{s+1}} - e^{-\beta|\sigma|(x_s-1)}) (1 - e^{-\beta|\sigma|})^{-1} & \text{if } \sigma < 0 \end{cases}
 \end{aligned}
 \tag{8}$$

5 Inverse Transform Sampling

Sampling from within a selected segment may be performed using various techniques, for example, using inverse transform sampling. In the following, inverse transform sampling is described for integer and continuous variables. Let x be a continuous random variable with values in the segment $x \in [x_L, x_U]$. With reference to the above discussion, x_L and x_U can be taken as segment bounds x_s, x_{s+1} , respectively. The partition function for a segment can be generically written as:

$$Z(\lambda) = \int_{x_L}^{x_U} dx \exp(-\lambda x) = \frac{1}{\lambda} [\exp(-\lambda x_L) - \exp(-\lambda x_U)]$$

The cumulative distribution function can be calculated as:

$$F(x) = \int_{x_L}^{x_U} dx' \exp(-\lambda x') = \frac{\exp(-\lambda x_L) - \exp(-\lambda x)}{\exp(-\lambda x_L) - \exp(-\lambda x_U)} = \frac{1 - \exp[-\lambda(x - x_L)]}{1 - \exp(-\lambda R)}$$

where R is defined as $R = x_U - x_L$. This formula can be inverted analytically, and setting $F(x) = y$ yields: $x = x_L - \frac{1}{\lambda} \ln[1 + y(e^{-\lambda R} - 1)]$ (9)

To sample a continuous variable within the segment $[x_L, x_U]$ with the distribution $\exp[-\lambda x]$, a variable y that is uniformly distributed in $[0,1]$ can be sampled randomly, and then equation (9) used.

The integer case may be sampled similarly, the cumulative distribution function is given by:

$$F(x) = \sum_{n=x_L}^x \exp[-\lambda n] = \begin{cases} \frac{1 - \exp[-\lambda(x - x_L + 1)]}{1 - \exp(-\lambda R)} & \lambda > 0 \\ \frac{1 - \exp[-\lambda(x - x_U - 1)]}{1 - \exp(-\lambda|R|)} & \lambda < 0 \end{cases} \tag{10}$$

Where here $R = x_U - x_L + 1$. The two cumulative distribution functions (9) and (10) are identical up to a displacement +1. In equation (10), argument x is an integer, so passing from a continuous random variable $y \in [0,1)$ requires the use of the floor and ceiling functions. Setting $F(x) = y$, the inverse of equation (10) yields:

$$x = \begin{cases} x_L + \left\lfloor -\frac{1}{\lambda} \ln[1 - y(1 - e^{-\lambda R})] \right\rfloor & \text{if } \lambda > 0, \\ x_U + \left\lceil \frac{1}{|\lambda|} \ln[1 - y(1 - e^{-|\lambda|R})] \right\rceil & \text{if } \lambda < 0. \end{cases}$$

The above-described techniques for integer variables will now be discussed with reference to Figure 3 and method 300. As discussed above, method 300 is an implementation of a method of operation of a computing system to update a sample in an optimization algorithm to improve convergency to feasibility. It will be understood that the above-described techniques for integer variables may similarly be applied to any of the methods discussed herein.

At 302, the processor receives a problem definition with a set of variables, an objective function defined over the set of variables, and one or more constraint functions, each of the constraint functions defined by at least one variable of the set of variables. In the present implementation, at least one variable of the set of variables is an integer variable.

At 304, sample values are received by the processor for the set of variables. This may be based on some known properties of the system, may be provided by another algorithm, may be randomly selected, or may be provided by other techniques as will be recognized by those of skill in the art. A value for a progress parameter, such as a temperature in the case of simulated annealing, is also received.

At 306, the processor selects a variable from the set of variables. For the purpose of this example, only integer variables will be discussed. However, this method may also be applied to continuous variables. Other variable types may be updated as discussed in further detail above. It will be understood that the order of acts 308 to 318 may be varied from the order shown in Figure 3, as discussed in further detail below.

At 308, the variable type is determined. In this example, the variable type is determined to be integer.

At 310, a sampling distribution is selected for the integer variables. As discussed above, there is not a single well-defined distribution that may be selected for integer variables. Instead, Gibbs sampling, referring to a technique for sampling from a conditional distribution, is performed. The sampling distribution is selected to be the conditional distribution given by the conditional partition function.

At 312, an objective energy change bias on the sample value for the variable and one or more terms of the objective function that include the variable is determined. As discussed above, for an integer variable this is calculated as the effective linear bias of the variable x_k , and is given by $\delta_{eff} = a_k + \sum_{i \neq k} w_{ik} x_i$.

At 314, a constraint energy change based on the sample value for the variable and each of the constraint functions defined by the variable is determined. This is given by $\delta_{eff}^{(c)} = \alpha_k^{(c)} + \sum_{i \neq k} w_{ik}^{(c)} x_i$.

At 318, an updated value is sampled for the variable based on the objective energy change and the constraint energy change as discussed above. The updated value may also be selected based on the current value of each constraint and the type of each constraint as above. In particular, if none of the constraints contribute to the linear bias for a given variable ($\delta_{eff}^{(c)} = 0$), then the new integer value is sampled from the partition function of the objective function, as given by equations (2) and (3) above. If the constraints do contribute to the linear bias, segments are created where the bias values from the constraints have the same value or are well defined. A segment is randomly chosen with a probability proportional to its partition function, and then a new integer value is randomly sampled from within that segment. The partition function of each segment is computed based on equation (8) above, and is based on the total energy change given by δ_{eff} and $\delta_{eff}^{(c)}$ and the progress parameter β .

At 320, it is determined if all of the variables have been considered. If there are other integer variables, the process above for sampling is repeated. If there are other types of variables, sampling is performed as described above.

At 322, an updated sample is returned, the updated sample comprising the updated value for each variable of the set of variables.

Continuous Variables

5 As discussed above, constrained problems may be defined over different variable types, such as binary, discrete, integer, and continuous. The variable types may determine the sampling distribution selected by the processor. In the above discussion of integer variables, there may not be a single well defined sampling distribution. Similarly, for continuous variables, there may not be a single well defined sampling distribution. Continuous
10 variables can take on any two particular real values, and can also take on all real values between those two values. For example, a continuous variable may take any value over the range of real numbers. Many types of optimization problems include decision variables that are represented by continuous variables. Some examples of such continuous variables include weight, volume, pressure, temperature, speed, flow rate, and elevation level.

15 One method of addressing continuous variables in optimization problems is using discretization of the continuous variables to represent the continuous variables using a set of integer or binary variables, and using the techniques described above for the redefined variable type. However, discretization of continuous variables may not provide a computationally efficient solution. For example, conversion of continuous variables into
20 integer variables may require an infinite upper bound to model continuous variables with arbitrary precision. Integer variables with infinite bounds may not be practical to implement in many applications. Using binary variables may similarly significantly increase the size of the problem, resulting in inefficient solving.

25 Inverse Transform Sampling for Continuous Variables

As mentioned above, the inverse transform sampling method for integer variables can be readily adapted for continuous variables, with the main differences being how the partition function is calculated and how the segments are split. The partition function for continuous variables can be written as:

$$\begin{aligned}
Z_\beta^k &= \int_{lb[k]}^{ub[k]} dx \exp\{-\beta f(x|\{x_i\}_{i \neq k})\} \\
&= \int_{lb[k]}^{ub[k]} dx \exp\{-\beta [A + \delta_{eff} x]\} \\
&= \exp\{-\beta A\} \frac{\exp\{-\beta \delta_{eff} lb[k]\} - \exp\{-\beta \delta_{eff} ub[k]\}}{\beta \delta_{eff}}
\end{aligned}$$

5 The prefactor $\exp\{-\beta A\}$ can be ignored and the partition function can be calculated analytically without the need to distinguish the two cases $\delta_{eff} > 0$ or $\delta_{eff} < 0$.

When considering a constraint g_c adjacent to variable x_k , this will split the domain $[lb[k], ub[k]]$ in two domains depending on where the binding value $\tilde{x}_k^{(c)} = -A_k^{(c)}/\delta_{eff}^{(c)}$. Here we briefly consider the case where $lb[k] \leq \tilde{x}_k^{(c)} \leq ub[k]$. In this case, if

10 $\delta_{eff}^{(c)} > 0$, the partition function is given by:

$$Z_\beta^k = \int_{lb[k]}^{\tilde{x}_k^{(c)}} dx \exp\{-\beta [\delta_{eff} x]\} + \int_{\tilde{x}_k^{(c)}}^{ub[k]} dx \exp\{-\beta [\delta_{eff} x + \lambda_c \delta_{eff}^{(c)} (x - \tilde{x}_k^{(c)})^\alpha]\}.$$

Comparing the expression above with the integer counter-part demonstrates

15 how sums are replaced by integrals and the binding values floor and ceiling functions are removed. When considering multiple constraints, the partition function is given by:

$$Z_\beta^{k,(s)} = \int_{x_s}^{x_{s+1}} dx \exp\left\{-\beta \left[\delta_{eff} x + \sum_{c \in C_s} \lambda_c \delta_{eff}^{(c)} (x - \tilde{x}_k^{(c)})^\alpha\right]\right\}.$$

20 Even in this case, we can use a uniform way to calculate the partition function for the L_0 and L_1 penalizations analytically. For the L_0 penalization we have:

$$Z_\beta^{k,(s)} = \exp\left\{-\beta \sum_{c \in C_s} \lambda_c \delta_{eff}^{(c)}\right\} \frac{\exp\{-\beta \delta_{eff} x_s\} - \exp\{-\beta \delta_{eff} x_{s+1}\}}{\beta \delta_{eff}} \quad (9)$$

25 while for the L_1 penalization we have:

$$Z_{\beta}^{k,(s)} = \exp \left\{ \beta \sum_{c \in C_s} \lambda_c \delta_{eff}^{(c)} \tilde{x}_k^{(c)} \right\} \left\{ \exp \left[-\beta \left(\delta_{eff} + \sum_{c \in C_s} \lambda_c \delta_{eff}^{(c)} \right) x_s \right] - \exp \left[-\beta \left(\delta_{eff} + \sum_{c \in C_s} \lambda_c \delta_{eff}^{(c)} \right) x_{s+1} \right] \right\} / \left\{ \beta \left[\delta_{eff} + \sum_{c \in C_s} \lambda_c \delta_{eff}^{(c)} \right] \right\} \quad (10)$$

5

The methods described above use inverse transform sampling or similar sampling methods over segments, and cannot be used for variables (integer or continuous) that have quadratic cases.

The cumulative distribution function can be inverted analytically only for cases having linear (x) or bi-linear ($x \times y$) terms. In the case of quadratic (x^2) terms, inverting the cumulative distribution function involves significant computational overhead such as the requirement to calculate the error function ($\text{erf}(x)$) or the imaginary error function ($\text{ierf}(x)$). Slice sampling may beneficially be performed in order to sample from the conditional distribution function without the requirement to invert the cumulative distribution function.

In some implementations, slice sampling may be used only for variables that are not included in other methods described herein. However, in other implementations, this method may be used to sample integer and continuous variables having linear, bi-linear, and quadratic interactions. This may also accommodate interactions between continuous variables and other types of variables.

Similar to the implementation discussed above, consider the case where there is a constrained quadratic model (CQM) with binary, integer, and continuous variables. Let $B \subseteq \{1, \dots, n\}$ denote the set of indices of the binary variables, that is, $x_i \in \{0, 1\}$ if $i \in B$. Similarly, let $C \subseteq \{1, \dots, n\}$ denote the set of indices of continuous variables, that is, $x_i \in \mathbb{R}$ if $i \in C$, and let $I \subseteq \{1, \dots, n\}$ denote the set of indices of integer variables, that is, $x_i \in \mathbb{Z}$ if $i \in I$.

25

The CQM assumes the form:

$$\begin{aligned} \text{Minimize:} \quad & \sum_i a_i x_i + \sum_{i \leq j} b_{ij} x_i x_j, \\ \text{Subject to:} \quad & \sum_i a_i^{(c)} x_i + \sum_{i \leq j} b_{ij}^{(c)} x_i x_j + c^{(c)} \leq 0, \\ & lb[k] \leq x_i \leq ub[k] \end{aligned}$$

30

The term $\sum_{i \leq j} b_{ij} x_i x_j$ includes the pure quadratic term $b_{ii} x_i^2$, although it is noted that for binary variables $x_i^2 = x_i$.

Consider the model without constraints, when updating the variable x_k ,
 5 where $k \in C$ or $k \in I$. In an implementation with an annealing solver and Gibbs sampling, the conditional probability distribution function of variable x_k given the values of the other variables $\{x_i\}_{i \neq k}$ can be written as $P(x_k | \{x_i\}_{i \neq k}) \propto \exp\{-\beta[b_{kk}x_k^2 + \delta_k x_k + A_k]\}$, where β is the inverse temperature at which a given increment of the optimization algorithm (e.g., annealing solver) is performed, δ_k is the effective linear bias of variables x_k and A_k is a
 10 constant factor that includes the constant d and the contribution of other variables $\{x_i\}_{i \neq k}$. β increases as the annealing proceeds.

If $b_{kk} = 0$ then the sampling for integer or continuous variables can be performed using the method discussed above for integer variables as there is no quadratic term. The cumulative distribution function of $p(x) = \exp(-\lambda x)$ can be inverted for both
 15 integer and continuous variables. When $b_{kk} \neq 0$, however, the function is no longer easily invertible. An implementation of sampling for quadratic terms and mixed variable interactions will be described below.

In the case of quadratic terms and mixed variable interactions, sampling for a given variable may be performed by slice sampling. A general description of slice sampling is
 20 provided in Neal, R., *Slice Sampling*, The Annals of Statistics, 2003, Vol. 31, No. 3, 705-767. Slice sampling generally is a Markov chain Monte Carlo algorithm for drawing random samples by sampling uniformly from the region under the graph of a variable's density function. For a given value of a variable x_0 , an auxiliary variable y is sampled uniformly between 0 and $f(x_0)$. A point (x, y) can be sampled from the line (or slice) at y where it is under the curve of $f(x_0)$.
 25 The value of x from the sampled point becomes the updated value for the variable x . While slice sampling is described below, it will be understood that in other implementations, inverse transform sampling may be used with the aid of the Newton-Raphson method, the bisection method, or other methods known in the art.

For a CQM problem, based on a current value of variable x_k , which can be
 30 denoted as x_k^0 , a value y can be sampled uniformly from between 0 and $P(x_k^0) =$

$\exp\left\{-\beta\left[b_{kk}(x_k^0)^2 + \delta_k x_k^0 + A_k\right]\right\}$, as given above. A new value for x_k can then be uniformly sampled from the region $\mathcal{X} = \{x_k: P(x_k) \geq y\}$.

Let $u \sim \text{RAND}(0,1)$ be a random number between 0 and 1. The condition $\exp\left\{-\beta\left[b_{kk}(x_k^0)^2 + \delta_k x_k^0 + A_k\right]\right\} \geq y$ can be written as $Ax_k^2 + Bx_k + C \leq 0$, where $A \equiv \beta b_{kk}$, $B \equiv \beta \delta_k$, $C = \beta A_k + \ln u + \ln P(x_k^0)$. These last two terms in the definition of C are provided by $\ln y \equiv \ln[u * P(x_k^0)] = \ln u + \ln P(x_k^0)$. Assuming x_k is a continuous variable, solving this inequality yields two values, $x_1 \leq x_2$ that are the zeros of the inequality.

There are two cases:

In the connected region where $x_k \in [x_1, x_2]$ with $x_1 \geq lb[k]$ and $x_2 \leq$
 10 $ub[k]$. The new value of x_k can be uniformly sampled from $[x_1, x_2]$.

In the disconnected region where $x_k \in [lb[k], x_1] \vee x_k \in [x_2, ub[k]]$, one of the two disconnected regions is selected with a probability proportional to $x_1 - lb[k]$ and $[x_2, ub[k]]$. A new value for x_k is uniformly sampled from the chosen region.

In the case where variable x_k is integer, the two values x_1, x_2 are replaced
 15 with $x_1, x_2 \rightarrow \lceil x_1 \rceil, \lfloor x_2 \rfloor$ in the connected region and $x_1, x_2 \rightarrow \lfloor x_1 \rfloor, \lceil x_2 \rceil$ in the disconnected region.

Turning now to constraints, “less than” inequalities will be discussed below. It will be understood that “greater than” equalities can be addressed similarly, and equality
 20 constraints can be addressed as a combination of two inequality constraints.

Considering the case of one constraint:

$$\sum_i a_i^{(c)} x_i + \sum_{i \leq j} b_{ij}^{(c)} x_i x_j + c^{(c)} \leq 0$$

As above, when updating variable x_k (which may be an integer or continuous variable), the other variables are fixed. The constraint can be written as:

$$\begin{aligned} & b_{kk}^{(c)} x_k^2 + \delta_k^{(c)} x_k + A_k^{(c)} \leq 0 \\ & \delta_k^{(c)} \equiv a_k^{(c)} + \sum_{i \neq k} b_{ik}^{(c)} x_i \\ & A_k^{(c)} \equiv \sum_{i, i \neq k} a_i^{(c)} x_i + \sum_{i, j \neq k} b_{ij}^{(c)} x_i x_j + c^{(c)} \end{aligned}$$

The solution of this inequality yields a maximum of two solutions, which can be denoted as x_1^c, x_2^c , with the assumption that $x_1^c < x_2^c$. These are also referred to as binding

values in the discussion above. The sign of $b_{kk}^{(c)}$ will determine if the region $[x_1^c, x_2^c]$ is feasible (that is, satisfies the constraints). See Figure 13, which shows an example implementation of a probability distribution function for a continuous variable.

Where $b_{kk} > 0$ the two solutions of the inequality are such that $lb[k] < x_1^c < x_2^c < ub[k]$. The inequality is satisfied in the segment $[x_1^c, x_2^c]$ (the parabola is below the axis). This is shown as region 1306 in Figure 13. Three segments can be defined: $[lb[k], x_1^c]$ where the inequality is not satisfied and the probability of sampling the variable x_k should be suppressed (region 1302 in Figure 13), $[x_1^c, x_2^c]$ where the inequality is satisfied and thus the probability of sampling x_k is not suppressed (region 1306 in Figure 13), and $[x_2^c, ub[k]]$ where the inequality is violated and the probability is suppressed (region 1304 in Figure 13).

Introducing a lagrangian $\lambda_c > 0$ for the constraint, the un-normalized probability distribution function can be written as:

$$P(x_k = n) \propto \begin{cases} \exp\left\{-\beta\left[b_{kk}n^2 + \delta_k n + A_k + \lambda_c\left(b_{kk}^{(c)}n^2 + \delta_k^{(c)}n + A_k^{(c)}\right)\right]\right\} & \text{if } n \in [lb[k], x_1^c] \vee n \in [x_2^c, ub[k]] \\ \exp\{-\beta[b_{kk}n^2 + \delta_k n + A_k]\} & \text{if } n \in [x_1^c, x_2^c] \end{cases}$$

The lagrangian coefficient (penalty parameters) may be provided to penalize infeasible solutions as discussed above in further detail, and in particular with reference to methods 800 and 900. The suppressed case can also be written as

$$P(x_k = n) \propto \exp\{-\beta[an^2 + bn + c]\}$$

With

$$a = b_{kk} + \lambda_c b_{kk}^{(c)}$$

$$b = \delta_k + \lambda_c \delta_k^{(c)}$$

$$c = A_k + \lambda_c A_k^{(c)}$$

The next value for variable x_k will then be sampled as follows: given the current value of the variable x_k^0 , the un-normalized probability distribution function is calculated and denoted as $P_0 = P(x_k^0)$. A slide y can be sampled uniformly in $y \sim U(0, P_0)$. For each of the segments the subregions S_i that are above the slice y are calculated, that is:

$$S_1 = \{x \in [lb[k], x_1^c] | P(x) \geq y\}$$

$$S_2 = \{x \in [x_1^c, x_2^c] | P(x) \geq y\}$$

$$S_3 = \{x \in [x_2^c, ub[k]] | P(x) \geq y\}$$

Each S_i can be empty (in case every x in the segment has $P(x) < y$) or disconnected, as above. Each of these subregions will be accompanied by an (un-normalized)

5 probability p_i that is the width of the region (for continuous variables) or the number of integers within the region (for integer variables). A sub-region will be selected with probability $\frac{p_i}{\sum_j p_j}$. The next value will be uniformly sampled from the selected region. Sampling from a selected segment can be performed by inverse transform sampling, or other sampling techniques as known in the art.

10 It will be understood that these principles can be extended to cases where there are no solutions to the equality above, with all of the domain suppressed accordingly, as well as to the case where there is only one solution ($x_1^c = x_2^c$). The above can also be generalized to cases where one or both of x_1^c, x_2^c are outside the variable domain $[\underline{x}_k, \bar{x}_k]$.

15 In the case of multiple constraints, the domain of variable x_k can be divided into a set of segments where some constraints are violated. After sampling a slice y , the method will proceed similarly to the implementation described above, with each segment being divided into sub-regions where $P(x) \geq y$. One of the sub regions is then sampled with a probability according to the total width of the sub region (or number of integers), and the new
20 value of variable x_k is sampled uniformly from within the selected sub-region.

L_2 norm penalization

In the discussions above with respect to integer variables, sampling for
25 constraint functions is described with respect to the L_0 and L_1 norms. The general partition function for a segment s is given by:

$$Z_\beta^{k,(s)} = \sum_{n=x_s}^{x_{s+1}} \exp \left\{ -\beta \left[\delta_{eff} n + \sum_{c \in C_s} \lambda_c \delta_{eff}^{(c)} (n - \hat{x}_k^{(c)})^\alpha \right] \right\}.$$

For the L_2 norm of the violation constraint, where $\alpha = 2$, inverse transform style sampling cannot be performed as the expression above will contain quadratic terms. However, slice sampling, as discussed above, can be used to accommodate the use of the L_2 norm for cases where the constraints have linear or bilinear terms. It will be understood that for constraints having quadratic terms the L_2 penalization would yield up to quartic terms, which is beyond the scope of the discussion below. Additional modification of the methods described herein may be required to accommodate quartic cases, or the L_0 and L_1 norms may be used in these cases.

Given a constraint with no quadratic terms, that is, $b_{kk}^c = 0$, the constraint for variable x_k can be rewritten as $A_k^c + \delta_k^c x_k \leq 0$. Assuming $\delta_k^c > 0$, the probability to sample $x_k > \tilde{x}_k^c = -A_k^c / \delta_k^c$ can be suppressed according to:

$$p(x_k = n > \tilde{x}_k^c) \propto \exp\{-\beta[\delta_k^c(n - \tilde{x}_k^c)^2]\}$$

The square in the exponential can be expanded as above. The slice sampling procedure described above can be used to sample a new value for x_k .

15

Pseudocode:

An example implementation of the methods described above will now be described.

Input: A CQM, a state x on the CQM, the current Lagrangian multiplier λ_c for each constraint c , the current energy value $E[c]$ for each constraint c , inverse temperature β , and the integer variable x_k to be updated.

The effective bias function is defined given a state, variable, and constraint as: $\delta_{eff,k}^{(c)} = a_{c,k} + \sum_{i \neq k} w_{ki,c} x_i$ (leaving out the constraint argument c just returns to the effective bias due to the main CQM objective). Note that in the definition of the effective bias the subscript k indicates that the quantity is recomputed for every variable x_k and for each adjacent constraint c .

In one example implementation, the following structures are defined:

- **‘QuadraticPenalty’**, with the properties ‘quadratic’, ‘linear’ and ‘offset’ (all real numbers) that encode the quantities introduced in the linear penalty equations

30

defined above. This is used to describe a quadratic penalty with quadratic and linear coefficients, where `offset` is the value of the penalty at $x_k = 0$.

The addition operation is also defined as: Given two `QuadraticPenalty`

5 objects QP1 and QP2, define QP = QP1+QP2 such that:

QP.quadratic = QP1.quadratic + QP2.quadratic

QP.linear = QP1.linear + QP2.linear

QP.offset = QP1.offset + QP2.offset.

- 10
- **`ConstraintPoint`**, with properties **`value`** (an integer), **`left`** and **`right`** (both **`QuadraticPenalty`**). This describes a point along the domain of x_i at which one or more penalties changes to a different quadratic function. The **`value`** refers to the value of x_k , and **`left`** and **`right`** refer to the different quadratic penalties (or sums of quadratic penalties) that become 0 closest to **`value`**. **`left`** refers to penalties coming from the negative direction of the domain of x_k which have negative **`slope`**, and

15

`right` refers to penalties going to the positive direction of the domain with positive **`slope`**. For two **`ConstraintPoint`** objects CP1 and CP2 the addition is defined only if $CP1.value == CP2.value$. If this is satisfied, $CP = CP1 + CP2$ is set with $CP.value = CP1.value$, $CP.left = CP1.left + CP2.left$, and $CP.right = CP1.right + CP2.right$.

20

 - **`PartialSegment`** with the properties **`lower\_bound`** (an integer) and **`quadratic\_penalty`** (a **`QuadraticPenalty`**). This corresponds to the starting point of each segment (segments as defined in the mathematical formulation section above) and the total quadratic penalty that is applied to the variable in the portion of the segment that follows the lower bound.

25

 - **`Segment`** with properties **`lower\_bound`**, **`upper\_bound`**, and "quadratic_penalty".

In one example implementation, the algorithm proceeds with:

Subroutine 1:

sample_integer_continuous_variable

Inputs:

- 30
- inverse temperature beta
 - current state x
 - variable xk to update

lagrangians list
 CQM model

5 Initialize 2 empty arrays of ConstraintPoint objects called tmp_constraints and constraints
 Initialize empty array of PartialSegment objects called partial_segments
 Initialize empty array of Segment objects called segments

FROM THE OBJECTIVE FUNCTION GET quadratic_bias, linear_bias

10 IF xk is an integer variable THEN:
 update_integer = 1
 ELSE:
 update_integer = 0

15 END IF

 num_active_constraints = 0
 FOR each constraint c adjacent to variable xk DO:
 lagrangian is the lagrangian parameter

20 tmp_constraints, num_active_constraints = PARSE_CONSTRAINT(c, lagrangian,
 update_integer, tmp_constraints, xk_lower, xk_upper, num_active_constraints)

 END FOR

25 SORT tmp_constraints by value and then by left and right penalty, quadratic first, offset
 last.
 constraints, num_constraints = COLLAPSE_CONSTRAINTS(tmp_constraints,
 num_active_constraints)
 partial_segments, num_partial_segments = MAKE_PARTIAL_SEGMENTS(quadratic_bias,
 30 linear_bias, lower_bound, upper_bound,
 collapsed_constraints, num_collapsed_constraints,
 partial_segments, update_integer)
 segments, num_segments = MAKE_SEGMENTS(partial_segments,
 num_partial_segments, lower_bound, upper_bound,
 35 segments, update_integer)

 new_value = SAMPLE_VALUE_FROM_SEGMENTS(segments, num_segments, beta)

 RETURN new_value

40 The next subroutine parses the constraint c , that is, extracts either the binding value $\tilde{x}_k^{(c)}$ or
 the two values x_1^c and x_2^c (the binding values) depending on if the constraint is linear or
 quadratic respectively and generates up to 2 constraint points:

45 **Subroutine 2:**

```

PARSE_CONSTRAINT
  Inputs:
    constraint c
    lagrangian
5    update_integer
    tmp_constraints
    num_active_constraints
    lower_bound
    upper_bound
10
  From the constraint c, get the quadratic coefficient b_kk, the effective linear bias delta_k and
  the offset A_k

    IF b_kk = 0 THEN
15      # linear constraint

      IF delta_k == 0 THEN
        skip this constraint
      END IF

20      xc = -delta_k / A_k
      if update_integer THEN
        IF delta_k > 0 THEN
          xc = CEIL(xc)
25        ELSE
          xc = FLOOR(xc)
        END IF
      END IF

30      IF using L0 penalty THEN
        QP = QUADRATIC_PENALTY(0, 0, lagrangian*abs(delta_k))
      ELSE IF using L1 penalty THEN
        QP = QUADRATIC_PENALTY(0, lagrangian*delta_k, -lagrangian*delta_k*xc)
      ELSE IF using L2 penalty THEN
35        QP = QUADRATIC_PENALTY(lagrangian*delta_k**2, -lagrangian*delta_k*A_k,
        lagrangian*A_k**2)
      END IF

      CP = ConstraintPoint(xc)
40      IF delta_k > 0 THEN
        CP.right = QP
      ELSE
        CP.left = QP
      END IF

45      tmp_constraints[num_active_constraints] = CP
      num_active_constraints ++

```

```

ELSE
  # The constraint is quadratic.
  # Here x1 and x2 are sorted if they are real
  x1, x2 = SOLVE_QUADRATIC_EQUATION(b_kk, delta_k, A_k)
5
  IF (x1 <= x2) AND (x1, x2 are real) THEN
    IF b_kk > 0 THEN

      IF update_integer THEN
10        x1 = CEIL(x1)
        x2 = FLOOR(x2)
      END IF

      IF using L0 penalty THEN
15        QP = QUADRATIC_PENALTY(0, 0, lagrangian)
      ELSE IF using L1 penalty THEN
        QP = QUADRATIC_PENALTY(lagrangian*b_kk, lagrangian*delta_k,
lagrangian*A_k)
      END IF # cannot use L2 penalty in this case
20

      CP = ConstraintPoint(x1)
      CP.left = QP
      tmp_constraints[num_active_constraints] = CP
      num_active_constraints ++
25

      CP = ConstraintPoint(x2)
      CP.right = QP
      tmp_constraints[num_active_constraints] = CP
      num_active_constraints ++
30

    ELSE IF b_kk < 0 THEN

      IF update_integer THEN
        x1 = FLOOR(x1)
35        x2 = CEIL(x2)
      END IF

      IF using L0 penalty THEN
        QP = QUADRATIC_PENALTY(0, 0, lagrangian)
40      ELSE IF using L1 penalty THEN
        QP = QUADRATIC_PENALTY(lagrangian*b_kk, lagrangian*delta_k,
lagrangian*A_k)
      END IF
45

      CP = ConstraintPoint(x1)
      CP.right = QP
      tmp_constraints[num_active_constraints] = CP
      num_active_constraints ++

```

```

        CP = ConstraintPoint(x2)
        CP.left = -QP
        tmp_constraints[num_active_constraints] = CP
5      num_active_constraints ++
      END IF

      ELSE IF x1, x2 are complex THEN
        IF b_kk > 0 THEN
10          # the constraint is always violated
          IF using L0 penalty THEN
            QP = QUADRATIC_PENALTY(0, 0, lagrangian)
          ELSE IF using L1 penalty THEN
            QP = QUADRATIC_PENALTY(lagrangian*b_kk, lagrangian*delta_k,
15      lagrangian*A_k)
          END IF
          CP = ConstraintPoint(lower_bound)
          CP.right = QP
          tmp_constraints[num_active_constraints] = CP
20      num_active_constraints ++
        END IF
      END IF
    END IF

25    RETURN tmp_constraints, num_active_constraints

```

The following subroutine scans the tmp_constraints and makes sure that if there are two or more constraint points within the same value, only one that is the sum of them remains:

```

30  Subroutine 3:
    COLLAPSE_CONSTRAINTS
    Inputs:
        tmp_constraints
        num_active_constraints
35    constraints

    Set num_collapsed_constraints = 0

    current_constraint = tmp_constraints[0]
40
    FOR c_i = 1, num_active_constraints-1 DO

        next_constraint = tmp_constraints[c_i]
        IF current_constraint.value == next_constraint.value THEN
45          current_constraint = current_constraint + next_constraint

```

```

        ELSE
            constraints[num_collapsed_constraints] = current_constraint
            current_constraint = next_constraint
            num_collapsed_constraint ++
5      END IF
    END FOR

    constraints[num_collapsed_constraints] = current_constraint
10   num_collapsed_constraint ++

    RETURN constraints, num_collapsed_constraints

```

15 The following subroutine creates the partial segments, that is, the points after which the conditional probability distribution function changes the coefficients.

Subroutine 4:

```

20  MAKE_PARTIAL_SEGMENTS
    Inputs:

        quadratic_bias
        linear_bias
25   lower_bound
        upper_bound
        collapsed_constraints
        num_collapsed_constraints
        partial_segments
30   update_integer

    // L1 case
    penalty_sum = QUADRATIC_PENALTY(quadratic_bias, linear_bias, 0);
35
    FOR i=0, num_collapsed_constraints-1 DO
        IF collapsed_constraints[i].value < lower_bound THEN
            penalty_sum += collapsed_constraints[i].right
        ELSE
40         penalty_sum += collapsed_constraints[i].left
        END IF
    END FOR

45   current_value = lower_bound
    num_partial_segments = 0
    FOR i=0, num_collapsed_constraints-1 DO
        IF collapsed_constraints[i].value < lower_bound THEN

```

```

        CONTINUE
    END IF

    IF collapsed_constraints[i].value > lower_bound + update_integer THEN
5      BREAK
    END IF

    IF collapsed_constraints[i].right is present THEN
        IF current_value != collapsed_constraints[i].value THEN
10      partial_segments[num_partial_segments].lower_bound = current_value
        partial_segments[num_partial_segments].quadratic_penalty = penalty_sum;
        num_partial_segments++;
        END IF

15      current_value = collapsed_constraints[i].value
        penalty_sum += collapsed_constraints[i].right

    END IF

20      IF collapsed_constraints[i].left is present THEN
        partial_segments[num_partial_segments].lower_bound = current_value
        partial_segments[num_partial_segments].quadratic_penalty = penalty_sum;
        current_value = constraints[i].value + update_integer;
        penalty_sum -= constraints[i].left;
25      num_partial_segments++;
    END IF

    END FOR

30      partial_segments[num_partial_segments].lower_bound = current_value
    partial_segments[num_partial_segments].quadratic_penalty = penalty_sum
    num_partial_segments++

    RETURN partial_segments, num_partial_segments
35

```

The following subroutine creates the segments from the array of partial segments created by the subroutine above:

Subroutine 5:

```

40  MAKE_SEGMENTS
    Inputs:
        partial_segments
        num_partial_segments
        lower_bound
45  upper_bound

```

```

segments
update_integer

5      num_segments = num_partial_segments - 1;
      segment_idx = 0;

      IF num_segments > 0 THEN

10         FOR i=0, num_partial_segments-2 DO
             IF partial_segments[i].lower_bound >= lower_bound THEN
                 segments[segment_idx].lower_bound = partial_segments[i].lower_bound
                 segments[segment_idx].upper_bound = partial_segments[i + 1].lower_bound -
update_integer
15         segments[segment_idx].quadratic_penalty =
partial_segments[i].quadratic_penalty
             segment_idx++
             ELSE
                 num_segments -= 1
20         END IF
             END FOR

             ELSE
25         # in case there are no partial segments
             num_segments++;
             segments[0].lower_bound = partial_segments[0].lower_bound;
             segments[0].upper_bound = partial_segments[0].lower_bound;
             segments[0].quadratic_penalty = partial_segments[0].quadratic_penalty;

30         END IF

             RETURN segments, num_segments

```

The final subroutine samples a value from a segment, and varies based on if the variable is
35 integer (6a) or continuous (6b).

Subroutine 6a:

```

SAMPLE_VALUE_FROM_SEGMENTS
40      Inputs:
          beta
          current_value
          segments
          num_segments
45      update_integer

```

```

FOR each segment DO
  IF update_integer THEN
    calculate the segment partition function with equation (8)
  ELSE
5    calculate the segment partition function with equations (9, 10)
  END IF
END FOR

10  SAMPLE one segment with probability proportional to its partition function
    SAMPLE new value using inverse transform sampling method on this segment
    RETURN new value

```

Subroutine 6b:

```

15  SAMPLE_VALUE_FROM_SEGMENTS
    Inputs:
    beta
    current_value
    segments
20  num_segments
    update_integer

    calculate log f0 from the segment that contains current_value
    SAMPLE u = RAND(0,1)
25
    FOR each segment DO
      calculate subregions where  $P(x) \geq f_0$ 
      set probability of each subregions proportional to its size
    END FOR
30
    SAMPLE one subregion randomly according to respective probability
    SAMPLE uniformly new value from the selected subregion
    RETURN new value

```

35 In an alternative implementation, sampling for continuous variables may be done from a linear programming model with the values for the variables of other types (i.e., discrete, integer, and binary) fixed. This may beneficially allow for solving a CQM having continuous variables without the addition of multiple variables to the problem, and may result in convergence of continuous variables towards an optimal solution at a similar rate to the

40 convergence of the other problem variables. When solving an optimization problem, where the variable type is continuous, sampling may be performed based on a linear programming model. Linear programming, also called linear optimization, is a mathematical model for

variables having linear relationships. Methods for solving linear programming problems are well known in the art.

After each sweep of an optimization algorithm such as simulated annealing, a linear programming problem can be created by fixing the values of the integer, binary, and discrete variables to their updated sampled values. A solution for the linear programming problem can then be found to define updated values for the continuous variables of the overall problem. The general structure of the linear problem does not change at each stage of the optimization algorithm, and thus there may be low overhead to solving the linear programming model at each stage. This may reduce the computational requirements of solving for continuous variables, as building a new linear programming model at each stage may introduce a large overhead, particularly where numerous iterations are required for convergence of the optimization algorithm.

In an implementation, optimization may proceed according to various implementations as discussed above for any integer, binary, or discrete variables in the problem. Updated values may also be sampled for variables having quadratic interactions, or interactions between continuous variables and other variable types, such as by slice sampling as discussed above. Once updated values for these variables have been sampled, the remaining continuous variables may have values sampled from a linear programming problem with the values for the other variables fixed at their updated values.

It will be understood that a linear programming model as described below does not support quadratic interactions between continuous variables, or interactions between continuous variables and other variable types. It will be further understood that in some implementations, a combination of sampling from the conditional probability distribution as discussed above, and sampling from a linear programming model may be used. For example, continuous variables having quadratic terms or interactions with other variable types may be sampled as discussed above using sampling from the conditional probability distribution through slice sampling. These continuous variables may then be fixed, along with the other variable types, and the remaining continuous variables may have values sampled from a linear programming model as discussed in further detail below.

Figure 14 is a flow diagram of an example method of operation of a computing system for finding a solution to a constrained problem or other problem that may be expressed as a constrained quadratic model, or for generating sample solutions that may be used as inputs to other algorithms, the problems having one or more continuous variables.

5 Figure 14 shows method 1400. Method 1400 may be executed on a hybrid computing system comprising at least one digital or classical processor and a quantum processor, for example hybrid computing system 100 of Figure 1, or by a classical computing system comprising at least one digital or classical processor.

Method 1400 comprises acts 1402 to 1444; however, a person skilled in the
10 art will understand that the number of acts illustrated is an example, and, in some implementations, certain acts may be omitted, further acts may be added, and/or the order of the acts may be changed.

Method 1400 starts at 1402, for example in response to a call or invocation from another routine or in response to an input by a user.

15 At 1402, the processor receives a problem definition. The problem definition includes a set of variables, an objective function defined over the set of variables, and one or more constraint functions, each of the constraint functions defined by at least one variable of the set of variables. The problem definition may be received from a user (e.g., entered via an input device), transmitted from another processor, retrieved from memory, or provided as the
20 output of another process performed by the processor. The objective function may be a quadratic function, and the problem definition may define a quadratic optimization problem. The set of variables may contain one or more of continuous, discrete, binary, and integer variables. The constraint functions may include one or more quadratic equality or inequality constraint functions.

25 In one implementation, the problem definition may be defined by:

$$\text{minimize } f = \sum_i a_i x_i + \sum_{i,j} b_{i,j} x_i x_j + \sum_l e_l y_l$$

with constraints defined by:

$$E_m = \sum_i a'_{i,m} x_i + \sum_{i,j} b'_{i,j,m} x_i x_j + \sum_l e'_{l,m} y_l + c_m \leq 0 \quad \forall m \in M$$

Where x_i are a set of discrete variables (which may include integer and binary variables, $x_i \in \mathbb{Z}$), a_i are linear coefficients for the discrete variables, $b_{i,j}$ are quadratic

- 5 coefficients between discrete variables, $a'_{i,m}$ are linear coefficients for discrete variables for constraint m , $b'_{i,j,m}$ are quadratic coefficients between discrete variables m , and c_m is a constant, as discussed above, and where y_l are continuous variables ($y_l \in \mathbb{R}$), e_l are linear biases for the continuous variables, $e'_{l,m}$ are linear coefficients for the continuous variables for constraint m , and E_m is the energy of constraint m .

- 10 The problem definition above does not include quadratic interactions between continuous variables and other variables. This model may beneficially be used to model mixed integer linear programming (MILP) problems, as well as some cases of mixed integer quadratic programming (MIQP) problems.

At 1404, the processor initializes a sample solution to the objective function.

- 15 The sample solution may be a random solution to the objective function. The random solution may be randomly selected from across the entire variable space, or the random solution may be selected within a range of the variable space based on known properties of the problem definition or other information. The sample solution may also be generated by another algorithm or provided as input by a user.

- 20 At 1406, the processor initializes a progress parameter. The progress parameter may be a set of incrementally changing values that define an optimization algorithm. For example, the progress parameter may be an inverse temperature, and the inverse temperature may increment from an initial high temperature to a final low temperature. In some implementations, an inverse temperature is provided as a progress
25 parameter for a simulated annealing algorithm. The selection of an inverse temperature may be performed as described above.

At 1408, the processor optionally initializes a penalty parameter. In some implementations the penalty parameter may be a Lagrange multiplier as discussed in further

detail below. In other implementations the penalty parameter may be selected as a constant value or a value that depends on one or more other variables.

At 1410, the processor initializes a linear programming problem for the continuous variables using the initial states of the variables defined at act 1404.

5

Considering the problem formulation above, given:

$$\text{minimize } f = \sum_i a_i x_i + \sum_{i,j} b_{i,j} x_i x_j + \sum_l e_l y_l$$

As above, the constraints may be restated as:

$$\sum_l e'_{l,m} y_l \leq - \sum_i a'_{i,m} x_i - \sum_{i,j} b'_{i,j,m} x_i x_j - c_m \quad \forall m \in M$$

10

The portions of the problem defined by the discrete, integer, and continuous variables may be set to a constant value (C_0, C_m), to provide a linear programming model:

$$\text{minimize } f = \sum_l e_l y_l + C_0$$

15

With the constraints:

$$\sum_l e'_{l,m} y_l \leq C_m \quad \forall m \in M$$

As discussed above, penalty parameters may also be included in the model, giving:

20

$$\text{minimize } f = \sum_l e_l y_l + \sum_m L_m v_m + C_0$$

With the constraints:

$$\sum_l e'_{l,m} y_l - v_m \leq C_m \quad \forall m \in M$$

25

Where L_m is a Lagrange multiplier, also referred to as a penalty parameter as discussed above, and v_m is the magnitude of the current violation for constraint m. The model may beneficially be built prior to the start of the optimization algorithm, and does not require reconstruction after each stage of the optimization. The general model structure (also called
 5 the model rim) does not change, beneficially allowing for the model to be updated and solved quickly. The part of the model that does change is the portion defined by the objective function and the fixed variables (the right-hand side of the equations above), which can be calculated after each update of these variables.

At 1412, the processor optionally calculates a current value of each constraint
 10 function at the sample values. For example, constraint functions may be given as $\sum A_{c,i} x_i + b_c = 0$, or as inequalities, depending on the constraints provided.

At 1414, the processor increments a stage of an optimization algorithm, such as by providing the progress parameter to the optimization algorithm. Optimization algorithms may include simulated annealing, parallel tempering, Markov Chain Monte Carlo
 15 techniques, branch and bound algorithms, and greedy algorithms, which may be performed by a classical computer. Optimization algorithms may also include algorithms performed by a quantum computer, such as quantum annealing, quantum approximate optimization algorithm (QAOA) or other noisy intermediate-scale quantum (NISQ) algorithms, quantum implemented fault-tolerant optimization methods, or other quantum optimization algorithms.
 20 Quantum computers may include quantum annealing processors or gate model based processors. On successive iterations the incremented optimization algorithm may provide samples.

At iteration k of the optimization algorithm (e.g., simulated annealing), sampling is performed for each integer, binary, and discrete variable as discussed below and in
 25 greater detail above, with the problem being considered with the values of the continuous variables fixed from the previous iteration ($k - 1$):

$$\text{minimize } f = \sum_i a_i x_i + \sum_{i,j} b_{i,j} x_i x_j + \sum_l e'_l y_l^{k-1}$$

With

$$\sum_i a'_{i,m} x_i + \sum_{i,j} b'_{i,j,m} x_i x_j + \sum_l e'_{l,m} y_m^{k-1} + c_m \leq 0 \quad \forall m \in M$$

$$x_i \in Z \text{ and } y_l \in R, y_l \geq 0$$

5 At 1416, the processor selects an i^{th} variable from a first set of variables of the set of variables. The first set of variables can include any integer, binary, or discrete variables, with a second set of variables including any continuous variables. The i^{th} variable has a current value from the sample initialized at act 1404.

10 At 1418, the processor determines a variable type for the variable. The variable type may be, for example, one of binary, discrete, and integer. It will be understood that a problem may contain a mixture of these types of variables, only one of these types of variables, and combinations thereof.

15 At 1420, the processor selects a sampling distribution based on the variable type. In some implementations, where the variable type is binary, the sampling distribution selected may be the Bernoulli distribution. In other implementations, where the variable type is discrete, the sampling distribution selected may be the SoftMax distribution. Alternatively, in some implementations, discrete variables may be included as binary variables using one-hot constraints and the sampling distribution may be the Bernoulli distribution. In other implementations, where the variable type is integer, sampling may be performed by Gibbs

20 sampling, that is, sampling on the conditional probability given the current state. Alternatively, Gibbs sampling may be performed on all variable types, with the conditional probability function being determined by the variable type. Integer variables may also be transformed to binary variables and sampling from the Bernoulli distribution may be performed. Sampling for integer and continuous variables may also be done by slice sampling as above.

25 At 1422, the processor determines an objective energy bias acting on the variable under consideration given the current values of all of the other variables. As discussed above, an optimization problem may be structured with an objective function that defines an energy, and during optimization a processor returns solutions with the goal of reducing this energy. In some implementations, such as where the variable under

30 consideration is a binary variable, an objective energy change (Δ_{E_i}) may be determined based

on the energy of the previous solution and the energy of the current sample value, with a decrease in energy suggesting an improved solution.

At 1424, the processor determines a constraint energy bias acting on the variable under consideration given the current values of all of the other variables and each of the constraint functions that involves the variable. A constraint energy bias may be defined by

$$\Delta_E = \sum_c (|\delta_{c,1}|^n - |\delta_{c,0}|^n)$$
for binary variables, as discussed above. The penalty applied to each constraint may be determined by the magnitude of the violation. In some implementations, the processor determines a total energy change based on the objective energy change and the constraint energy change. The constraint functions are indirectly included in this energy by the processor by the addition of an energy term considering the constraints.

At 1426, the processor samples an updated value for the variable from the sampling distribution based on the objective and constraint energy biases and the progress parameter as discussed in further detail above.

At 1428, the processor evaluates if all of the variables of the first set of variables have been considered. If all of the variables have not been considered, control passes back to act 1416, and the next variable is selected. Once all of the variables have been considered, control passes to act 1430. In some implementations, the processor may incrementally consider each variable in turn of the first set of variables and evaluate that all of the variables of the first set of variables have been considered when the last variable of the first set of variables has been considered.

At 1430, the processor fixes the values of the first set of variables based on their updated values from above.

At 1432, the processor solves the linear programming problem for the continuous variables in the second set of variables and samples updated values for the continuous variables. The linear programming problem can be solved using any technique known in the art, such as using an interior-point method or a simplex algorithm.

When considering the continuous variables, sampling is performed with the values for the integer, binary, and discrete variables fixed. That is, using a given integer/binary state at iteration k , that is, variable x_i^k , the following linear programming problem is used to sample values for the continuous variables:

$$\text{minimize } f = \sum_i a_i x_i^k + \sum_{i,j} b_{i,j} x_i^k x_j^k + \sum_l e_l y_l + u_m q_m$$

$$\sum_i a'_{i,m} x_i^k + \sum_{i,j} b'_{i,j,m} x_i^k x_j^k + \sum_k e'_{k,m} y_{k,m} - q_m + c_m \leq 0 \quad \forall m \in M$$

5 In the equations above, q_m is a positive continuous variable and u_m is a Lagrange multiplier. q_m is added to the left-hand side of each constraint to avoid infeasibility in the linear programming problem. q_m is then penalized and added to the objective function at the current value of the Lagrange multiplier. The linear programming solver defines the value of q_m alone with the problem continuous variables. This relaxation technique avoids
10 infeasibility, but is penalized.

At each stage of the optimization algorithm, the structure of the linear programming problem does not change, but the right-hand side of the constraints and the coefficient for the variable q_m are updated.

For large scale problems the linear programming problem may beneficially be
15 split into multiple smaller linear programming problems (subproblems) where the continuous variables are randomly assigned to one of the subproblems. The assignment may also be optimized to reduce the number of interactions between variables in different subproblems. When sampling for a variable in a given subproblem, the values of the continuous variables in all other subproblems are fixed at their most recent value. This may beneficially allow for
20 convergence of continuous variables to near optimum or global optimum points at the same rate as other variables.

For example, given G as the set of continuous variables, the variables may be randomly split into N groups with equal sizes $G_n = \{x_i | i \text{ in } G_n\}$. Variables and constraints can be defined for each group, and a linear programming problem can be built for each variable
25 and constraint group. Each linear programming problem may then be solved sequentially, with the related variables being updated.

At 1434, the processor updates the states for the continuous variables based on the results of the linear programming problem.

At 1436, the processor updates the energy biases for the objective and constraint functions based on the sampled values for the continuous variables.

At 1438, the processor increments the progress parameter, e.g., the temperature for simulated annealing.

5 At 1440, optionally, the penalty parameter for each constraint may be adjusted. In some implementations the penalty parameter may be adjusted based on the change in energy for the constraint function defined by the i^{th} variable and the progress parameter. In other implementations, the penalty parameter may be adjusted as described in methods 800 and 900, discussed in further detail below.

10 At 1442, the processor evaluates one or more termination criteria. In some implementations, the termination criteria may be a value of the progress parameter. In other implementations, the termination criteria may include a number of iterations, an amount of time, a threshold average change in energy between updates, a measure of the quality of the current values for the variables, or other metrics as are known in the art. If the termination
15 criteria are not met, the method continues with act 1414. As discussed above, incrementing a stage of an optimization algorithm for the objective function may include incrementing a simulated annealing or parallel tempering algorithm. In other implementations the optimization algorithm may be a MCMC algorithm or a greedy algorithm.

 If the one or more termination criteria has been met, control passes to 1444,
20 where a solution is output. At 1444, method 1400 terminates, until it is, for example, invoked again. The solution output at act 1444 may be passed to other algorithms, such as a quantum annealing algorithm.

 In some implementations, after outputting a solution at act 1444, method 1400 may begin again with a new sample being initialized at act 1404. In some
25 implementations, method 1400 may be performed multiple times in parallel starting from different initialized samples or a series of randomly generated samples. In some implementations, the solutions may be paired, and binary problems may be built to evaluate the series of solutions using Cross-Boltzmann updates.

An example implementation is described in the following pseudocode, which may be combined with the pseudocode discussed above.

Pseudocode:

5 For iteration $k = 0$

 initialize β, L^k, x^k, y^k

x initialized randomly, y is initialized to zero

 calculate $E_m^k = \text{compute_energy}(x^k, y^k)$

 While sweeping:

10 $E_m^{k+1}, x^{k+1} \leftarrow \text{sample discrete}$

 update energy:

$$E_m^{k+1} = \sum_l e'_{l,m} y_l^k + \sum_i a'_{i,m} x_i^{k+1} + \sum_{i,j} b'_{i,j,m} x_i^k x_j^{k+1} + c_m$$

$E_m^{k+1}, y^{k+1} \leftarrow \text{solve for continuous and update energy}$

$$rhs_m = E_m^{k+1} - \sum_l e'_{l,m} y_l^k$$

15 update LP objective with L^k

$y_l^{k+1} \leftarrow \text{solve LP}(rhs_m, L^k)$

$$E^{k+1} = E^{k+1} + \sum_l e'_{l,m} y_l^k - \sum_l e'_{l,m} y_l^{k+1} = rhs_m + \sum_l e_l y_l^{k+1}$$

20 The above described method(s), process(es), or technique(s) could be implemented by a series of processor readable instructions stored on one or more nontransitory processor-readable media. Some examples of the above described method(s), process(es), or technique(s) method are performed in part by a specialized device such as an adiabatic quantum computer or a quantum annealer or a system to program or otherwise control operation of an adiabatic quantum computer or a quantum annealer, for instance a

computer that includes at least one digital processor. The above described method(s), process(es), or technique(s) may include various acts, though those of skill in the art will appreciate that in alternative examples certain acts may be omitted and/or additional acts may be added. Those of skill in the art will appreciate that the illustrated order of the acts is shown for exemplary purposes only and may change in alternative examples. Some of the exemplary acts or operations of the above described method(s), process(es), or technique(s) are performed iteratively. Some acts of the above described method(s), process(es), or technique(s) can be performed during each iteration, after a plurality of iterations, or at the end of all the iterations.

10 The above description of illustrated implementations, including what is described in the Abstract, is not intended to be exhaustive or to limit the implementations to the precise forms disclosed. Although specific implementations of and examples are described herein for illustrative purposes, various equivalent modifications can be made without departing from the spirit and scope of the disclosure, as will be recognized by those skilled in the relevant art. The teachings provided herein of the various implementations can be applied to other methods of quantum computation, not necessarily the exemplary methods for quantum computation generally described above.

20 The various implementations described above can be combined to provide further implementations. All of the commonly assigned US patent application publications, US patent applications, foreign patents, and foreign patent applications referred to in this specification and/or listed in the Application Data Sheet are incorporated herein by reference, in their entirety, including but not limited to:

 U.S. Provisional Patent Application No. 62/951,749;
 U.S. Provisional Patent Application No. 63/174,097;
25 U.S. Provisional Patent Application No. 63/250,466;
 U.S. Patent Application Publication Nos. 2014/0344322 and 2020/0234172;
and

U.S. Patent Nos. 7,533,068; 8,008,942; 8,195,596; 8,190,548; and 8,421,053.

These and other changes can be made to the implementations in light of the above-detailed description. In general, in the following claims, the terms used should not be construed to limit the claims to the specific implementations disclosed in the specification and
5 the claims, but should be construed to include all possible implementations along with the full scope of equivalents to which such claims are entitled. Accordingly, the claims are not limited by the disclosure.

CLAIMS

1. A method of operation of a computing system to update a sample in an optimization algorithm to improve convergence to feasibility, the method being performed by a processor, the method comprising:

receiving a problem definition comprising a set of variables, an objective function defined over the set of variables, and one or more constraint functions, each of the constraint functions defined by at least one variable of the set of variables;

receiving sample values for the set of variables and a value for a progress parameter;

for each variable of the set of variables:

determining a variable type for the variable;

selecting a sampling distribution based on the variable type;

determining an objective energy bias based on the sample value for the variable and one or more terms of the objective function that include the variable;

determining one or more constraint energy biases based on the sample value for the variable and each of the constraint functions defined by the variable; and

sampling an updated value for the variable from the sampling distribution based on the objective energy bias, the one or more constraint energy biases, and the progress parameter; and

returning an updated sample, the updated sample comprising the updated value for each variable of the set of variables.

2. The method of claim 1, further comprising: receiving a value for a penalty parameter, and wherein sampling an updated value for the variable from the sampling distribution further comprises sampling an updated value for the variable from the sampling distribution based on the value for a penalty parameter.

3. The method of claim 2, wherein receiving a value for the penalty parameter comprises receiving a value for a Lagrange parameter that depends on the value for the progress parameter.

4. The method of any one of claims 1 through 3, wherein determining the variable type for the variable comprises determining that the variable type is one of binary, discrete, integer, or continuous.
5. The method of claim 4, wherein determining that the variable type is one of binary, discrete, integer, or continuous comprises determining that the variable type is binary, and wherein selecting a sampling distribution based on the variable type comprises selecting a Bernoulli distribution.
6. The method of claim 4, wherein determining that the variable type is one of binary, discrete, integer, or continuous comprises determining that the variable type is discrete, and wherein selecting a sampling distribution based on the variable type comprises selecting a SoftMax distribution.
7. The method of claim 4, wherein determining that the variable type is one of binary, discrete, integer, or continuous comprises determining that the variable type is one of integer or continuous, and wherein selecting a sampling distribution based on the variable type comprises selecting a conditional probability distribution.
8. The method of claim 7, wherein sampling an updated value for the variable from the sampling distribution comprises slice sampling from the conditional probability distribution.
9. The method of any one of claims 1 through 8, wherein receiving the value for the progress parameter comprises receiving an inverse temperature.
10. A system for updating a sample in an optimization algorithm, the system comprising:
 - at least one non-transitory processor-readable medium that stores at least one of processor executable instructions and data; and
 - at least one processor communicatively coupled to the least one non-transitory processor-readable medium, which, in response to execution of the at least one of processor executable instructions and data performs the method of any of claims 1 through 9.

11. A method of operation of a computing system, the computing system comprising one or more processors, the method being performed by at least one of the one or more processors, the method comprising:

receiving a problem definition comprising a set of variables, an objective function defined over the set of variables, and one or more constraint functions, each of the constraint functions defined by at least one variable of the set of variables;

initializing a sample solution to the objective function and a progress parameter;

iteratively until a termination criteria is met:

incrementing a stage of an optimization algorithm;

for each variable in the set of variables:

selecting an i^{th} variable from the set of variables, the i^{th} variable having a current value;

calculating an objective energy bias for the objective function based on the current value of the i^{th} variable;

calculating a constraint energy bias for each constraint function defined by the i^{th} variable based on the current value of the i^{th} variable; and

sampling an updated value for the i^{th} variable based on the objective energy bias and the constraint energy bias, the updated value replacing the current value;

incrementing a progress parameter;

evaluating a termination criteria; and

outputting a solution comprising the current values for the set of variables.

12. The method of claim 11, wherein:

selecting an i^{th} variable from the set of variables comprises selecting a binary variable, the binary variable having a current value and an alternative value;

calculating an objective energy bias for the objective function based on the current value of the i^{th} variable comprises calculating a difference in energy for the objective function between the current value and the alternative value for the binary variable;

calculating a constraint energy bias based on each constraint function defined by the i^{th} variable based on the current value of the i^{th} variable comprises calculating a difference in

energy for each constraint function defined by the binary variable between the current value of the binary variable and the alternative value for the binary variable; and

sampling an updated value for the i^{th} variable based on the objective energy bias and the constraint energy bias comprises sampling an updated value for the i^{th} variable based on the difference in energy values for the objective function and each constraint function defined by the i^{th} variable.

13. The method of claim 12, further comprising:

initializing a penalty parameter; and

adjusting the penalty parameter for each constraint function based on the difference in energy for the constraint function defined by the i^{th} variable and the progress parameter; and

wherein calculating the difference in energy for each constraint function defined by the i^{th} variable includes penalizing each constraint function by the penalty parameter.

14. The method of any one of claims 11 through 13, wherein incrementing a stage of an optimization algorithm for the objective function includes incrementing one of a simulated annealing or a parallel tempering algorithm.

15. The method of any one of claims 11 through 14, wherein receiving a problem definition comprising an objective function and one or more constraint functions comprises receiving a problem definition comprising a quadratic objective function and one or more quadratic equality or inequality constraint functions.

16. The method of any one of claims 11, 14, and 15, wherein receiving a problem definition comprising a set of variables comprises receiving a problem definition comprising a set of one or more of binary, integer, or discrete variables.

17. The method of claim 16, wherein receiving a problem definition comprising a set of variables comprises receiving a problem definition comprising one or more integer variables, and wherein sampling an updated value for the i^{th} variable comprises performing sampling

from a conditional probability distribution, the i^{th} variable comprising an integer variable from the one or more integer variables.

18. The method of any one of claims 11 through 17, wherein the termination criteria comprises one of a number of iterations, an amount of time, an average change in value limit, or a value of the progress parameter.

19. A system for use in optimization, the system comprising:
 at least one non-transitory processor-readable medium that stores at least one of processor executable instructions and data; and
 at least one processor communicatively coupled to the least one non-transitory processor-readable medium, which, in response to execution of the at least one of processor executable instructions and data, performs the method of any of claims 11 through 18.

20. The system of claim 19, further comprising a quantum processor, and wherein, after performing the method of any of claims 11 through 18, the at least one processor instructs the quantum processor to perform quantum annealing based on the outputted solution.

21. A method of operation of a hybrid computing system, the hybrid computing system comprising a quantum processor and a classical processor, the method being performed by the classical processor, the method comprising:

receiving a constrained quadratic optimization problem, the constrained quadratic optimization problem comprising a set of variables, an objective function defined over the set of variables, one or more constraint functions, each of the constraint functions defined by at least one variable of the set of variables, and a progress parameter for the optimization, the progress parameter comprising a set of values incrementing between an initial value and a final value;

iteratively until the final value of the progress parameter is reached:

sampling a sample set of values for the set of variables from an optimization algorithm;

updating the sample set of values with an update algorithm comprising:

for each variable of the set of variables:

determining a variable type for the variable;

selecting a sampling distribution based on the variable type;
 determining an objective energy bias based on a sample value for the variable from the sample set of values and one or more terms of the objective function that include the variable;
 determining one or more constraint energy bias based on the sample value for the variable and each of the constraint functions defined by the variable; and
 sampling an updated value for the variable from the sampling distribution based on the objective energy bias, the one or more constraint energy biases, and the progress parameter; and
 returning an updated sample, the updated sample comprising the updated value for each variable of the set of variables;
 incrementing the progress parameter;
 transmitting one or more final samples to a quantum processor;
 instructing the quantum processor to refine the samples; and
 outputting solutions.

22. The method of claim 21, wherein transmitting one or more final samples to a quantum processor comprises transmitting pairs of samples to the quantum processor, and wherein instructing the quantum processor to refine the samples comprises instructing the quantum processor to perform quantum annealing to select between the samples.

23. The method of any one of claims 21 and 22, further comprising: returning the outputted solutions as a sample set of values for the set of variables as an input to the optimization algorithm.

24. A hybrid computing system, the hybrid computing system comprising:
 a quantum processor and a classical processor;
 at least one non-transitory processor-readable medium that stores at least one of processor executable instructions and data; and
 at least one processor communicatively coupled to the least one non-transitory processor-readable medium, which, in response to execution of the at least one of processor executable instructions and data, performs the method of any of claims 21 through 23.

25. A method of operation of a computing system to direct a search space towards feasibility to improve performance of the computing system, the computing system comprising one or more processors, the method being performed by at least one of the one or more processors, the method comprising:

- receiving a sample from an optimization;
- determining an energy value for one or more constraint functions;
- evaluating feasibility of the sample;
- if the sample is not feasible, increasing a penalty value;
- if the sample is feasible, decreasing a penalty value; and
- returning a penalty value to an optimization algorithm.

26. The method of claim 25, further comprising: determining if a violation has been reduced in comparison with a previous sample and increasing an initial adjuster value if the violation has not been reduced.

27. The method of any one of claims 25 and 26, further comprising: determining if a current best solution has been improved in comparison with a previous sample and increasing an initial adjuster value if the current best solution has not been improved.

28. A method of operation of a computing system, the computing system comprising one or more processors, the method being performed by at least one of the one or more processors, the method comprising:

- receiving a problem definition comprising a set of variables, an objective function defined over the set of variables, and one or more constraint functions, each of the constraint functions defined by at least one variable of the set of variables, the set of variables comprising a first subset of variables having at least one variable that is one of binary, integer, discrete, or continuous and a second subset of variables having at least one variable that is continuous;
- initializing a sample solution to the objective function and a progress parameter;
- initializing a continuous problem defined over the second subset of variables, the continuous problem comprising a linear programming model;
- iteratively until a termination criteria is met:

incrementing a stage of an optimization algorithm;
for each variable in the first subset of variables:
 determining a variable type for the variable;
 selecting a sampling distribution based on the variable type;
 determining an objective energy bias based on the sample value for
the variable and one or more terms of the objective function that include the variable;
 determining one or more constraint energy biases based on the
sample value for the variable and each of the constraint functions defined by the variable;
 sampling an updated value for the variable from the sampling
distribution based on the objective energy bias, the one or more constraint energy biases, and
the progress parameter; and
 returning an updated sample, the updated sample comprising the
updated value for each variable of the set of variables;
solving the linear programming model for the second subset of variables with
the values for each variable in the first subset of variables fixed at the updated value;
sampling an updated value for each variable in the second subset of variables
based on the solved linear programming model, the updated value replacing a current value;
updating the objective energy bias and the one or more constraint energy
biases based on the updated values for the second subset of variables;
incrementing the progress parameter;
evaluating a termination criteria; and
outputting a solution comprising the updated values for the set of variables.

29. The method of claim 28, further comprising:

 initializing one or more penalty parameters;
 wherein sampling an updated value for the variable from the sampling
distribution further comprises sampling an updated value for the variable from the sampling
distribution based on at least one value for the one or more penalty parameters; and
 wherein sampling an updated value for each variable in the second subset of
variables comprises sampling an updated value for each variable in the second subset of
variables based on the solved linear programming model and at least one value for the one or
more penalty parameters.

30. The method of claim 29, wherein initializing one or more penalty parameters comprises initializing at least one Lagrange parameter that depends on the value for the progress parameter.
31. The method of any one of claims 28 through 30, wherein determining the variable type for the variable comprises determining that the variable type is one of binary, discrete, integer, or continuous.
32. The method of claim 31, wherein determining that the variable type is one of binary, discrete, integer, or continuous comprises determining that the variable type is binary, and wherein selecting a sampling distribution based on the variable type comprises selecting a Bernoulli distribution.
33. The method of claim 31, wherein determining that the variable type is one of binary, discrete, integer, or continuous comprises determining that the variable type is discrete, and wherein selecting a sampling distribution based on the variable type comprises selecting a SoftMax distribution.
34. The method of claim 31, wherein determining that the variable type is one of binary, discrete, integer, or continuous comprises determining that the variable type is one of integer or continuous, and wherein selecting a sampling distribution based on the variable type comprises selecting a conditional probability distribution.
35. The method of claim 34, wherein sampling an updated value for the variable from the sampling distribution comprises slice sampling from the conditional probability distribution.
36. The method of any one of claims 28 through 35, wherein receiving the value for the progress parameter comprises receiving an inverse temperature.
37. The method of any one of claims 28 through 36, wherein incrementing a stage of an optimization algorithm for the objective function includes incrementing one of a simulated annealing or a parallel tempering algorithm.

38. The method of any one of claims 28 through 37, wherein receiving a problem definition comprising an objective function and one or more constraint functions comprises receiving a problem definition comprising a quadratic objective function and one or more quadratic equality or inequality constraint functions.

39. The method of any one of claims 28 through 38, wherein the termination criteria comprises one of a number of iterations, an amount of time, an average change in value limit, or a value of the progress parameter.

40. A system for use in optimization, the system comprising:
at least one non-transitory processor-readable medium that stores at least one of processor executable instructions and data; and
at least one processor communicatively coupled to the least one non-transitory processor-readable medium, which, in response to execution of the at least one of processor executable instructions and data, performs the method of any of claims 28 through 39.

41. The system of claim 40, further comprising a quantum processor, and wherein, after performing the method of any of claims 28 through 39, the at least one processor instructs the quantum processor to perform quantum annealing based on the outputted solution.

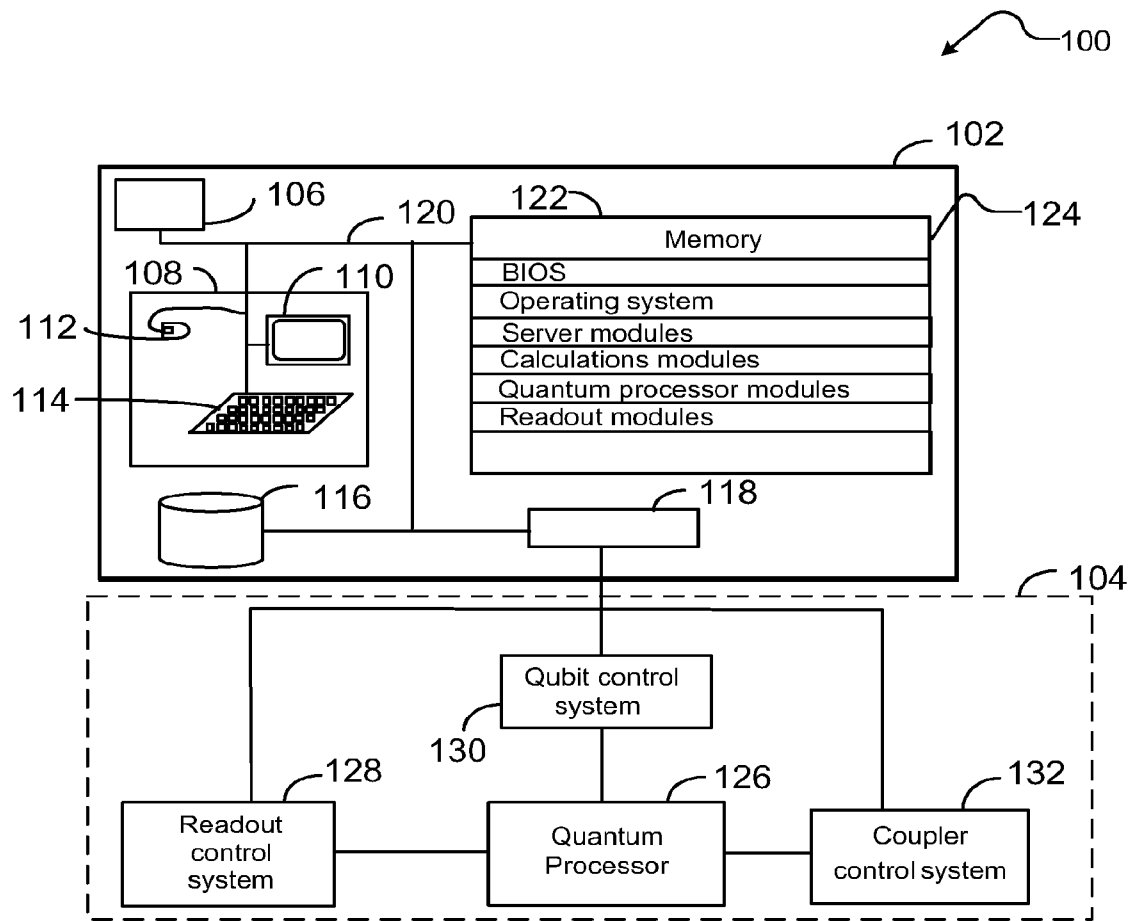


FIGURE 1

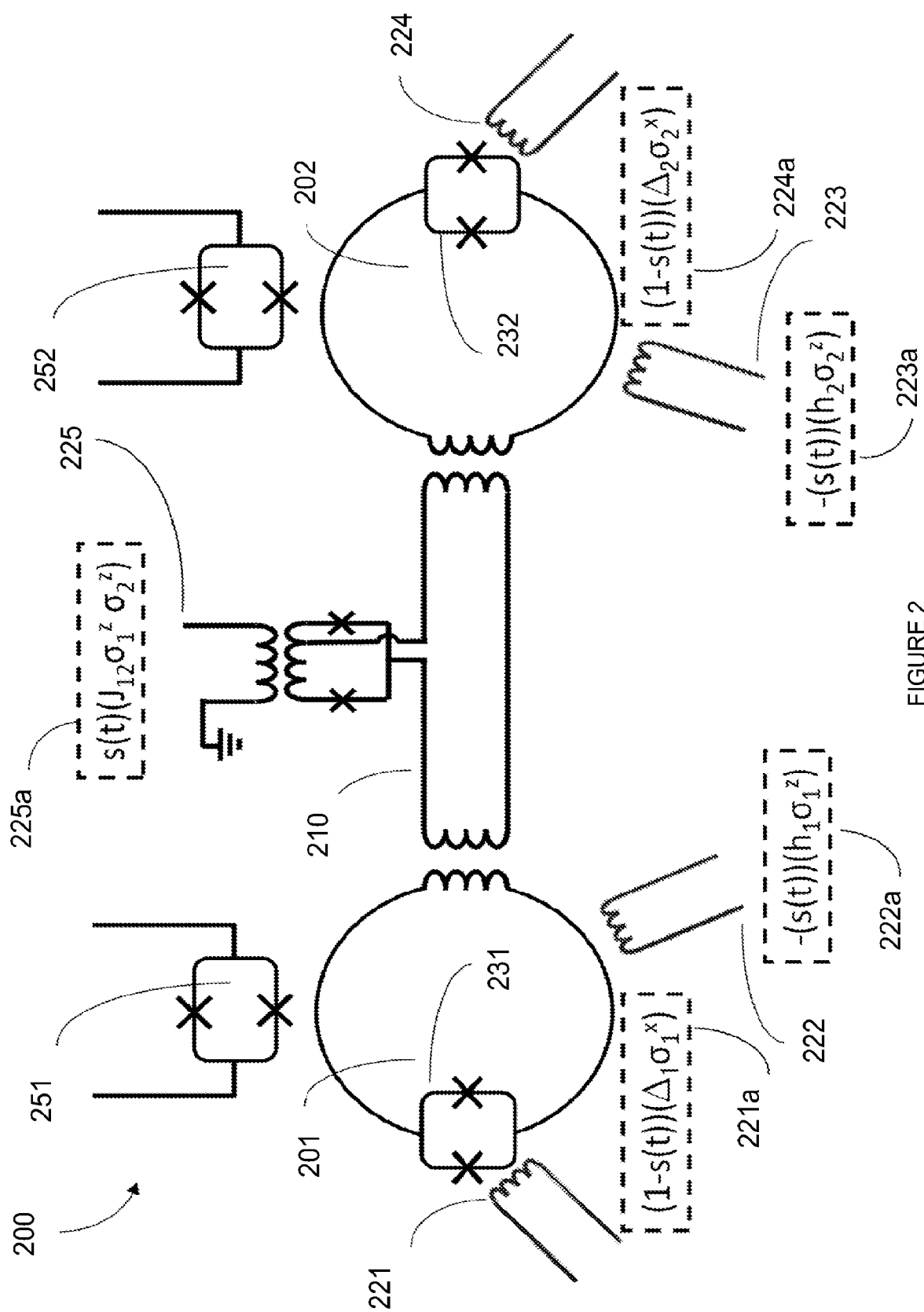


FIGURE 2

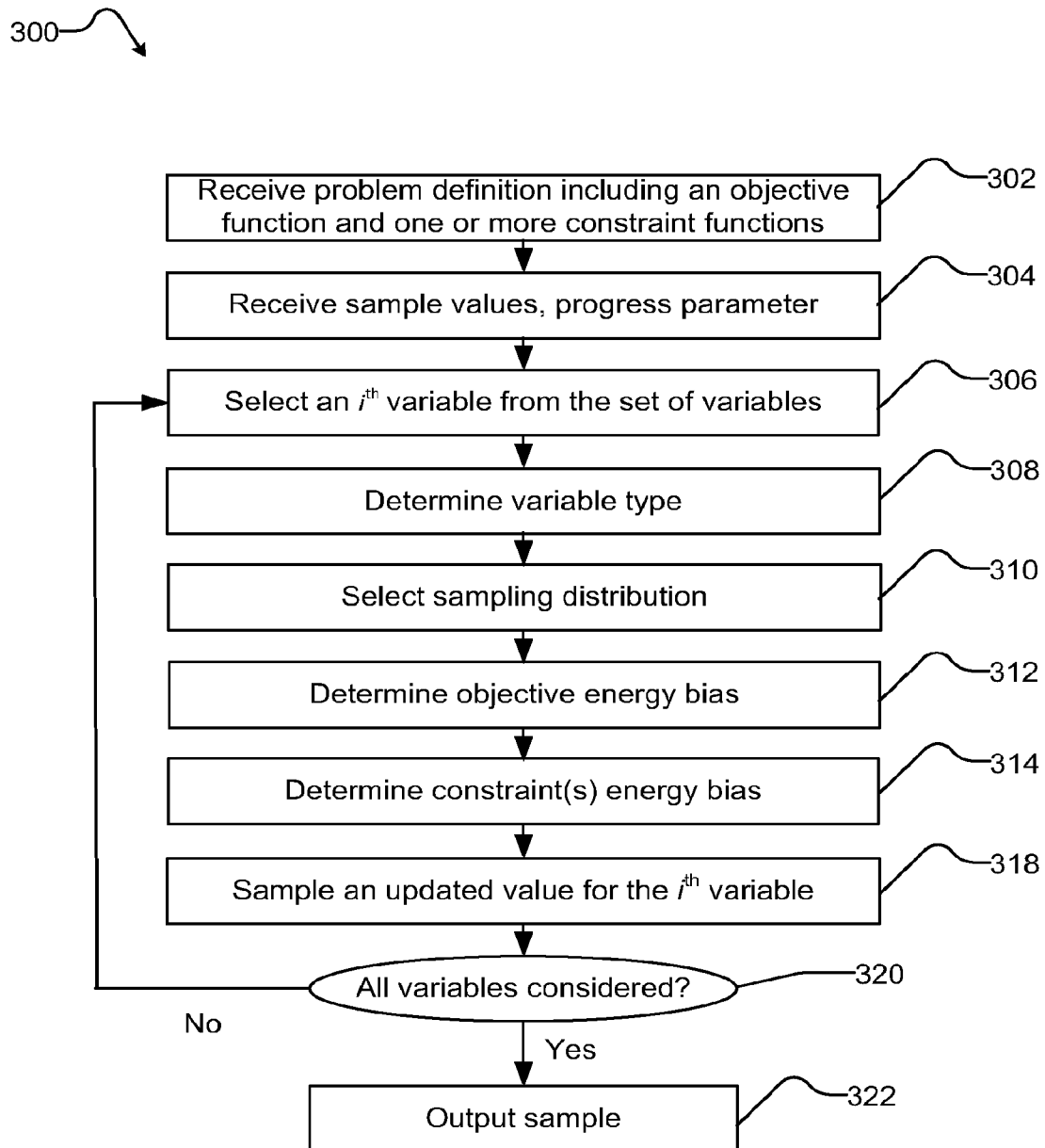


FIGURE 3

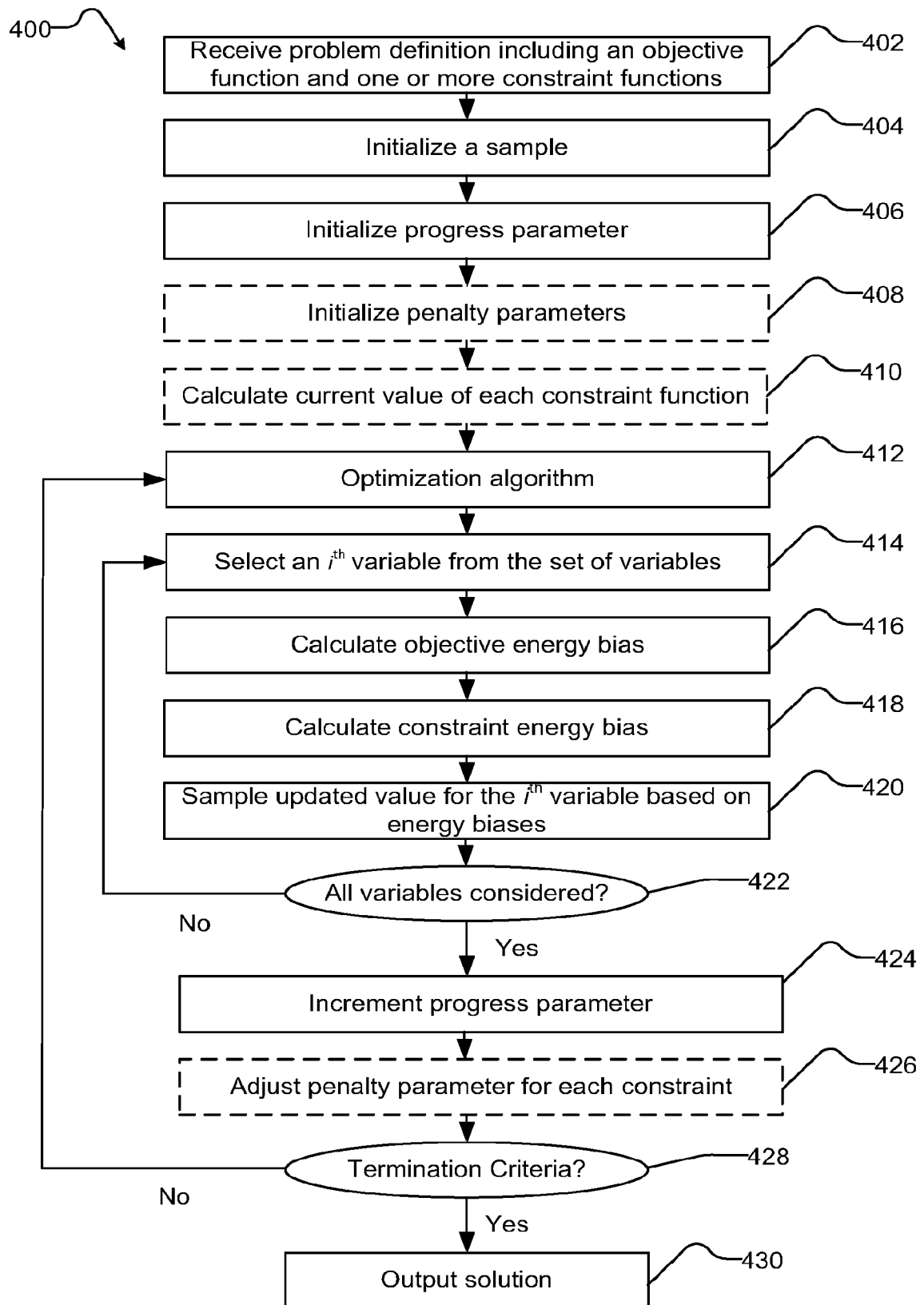


FIGURE 4

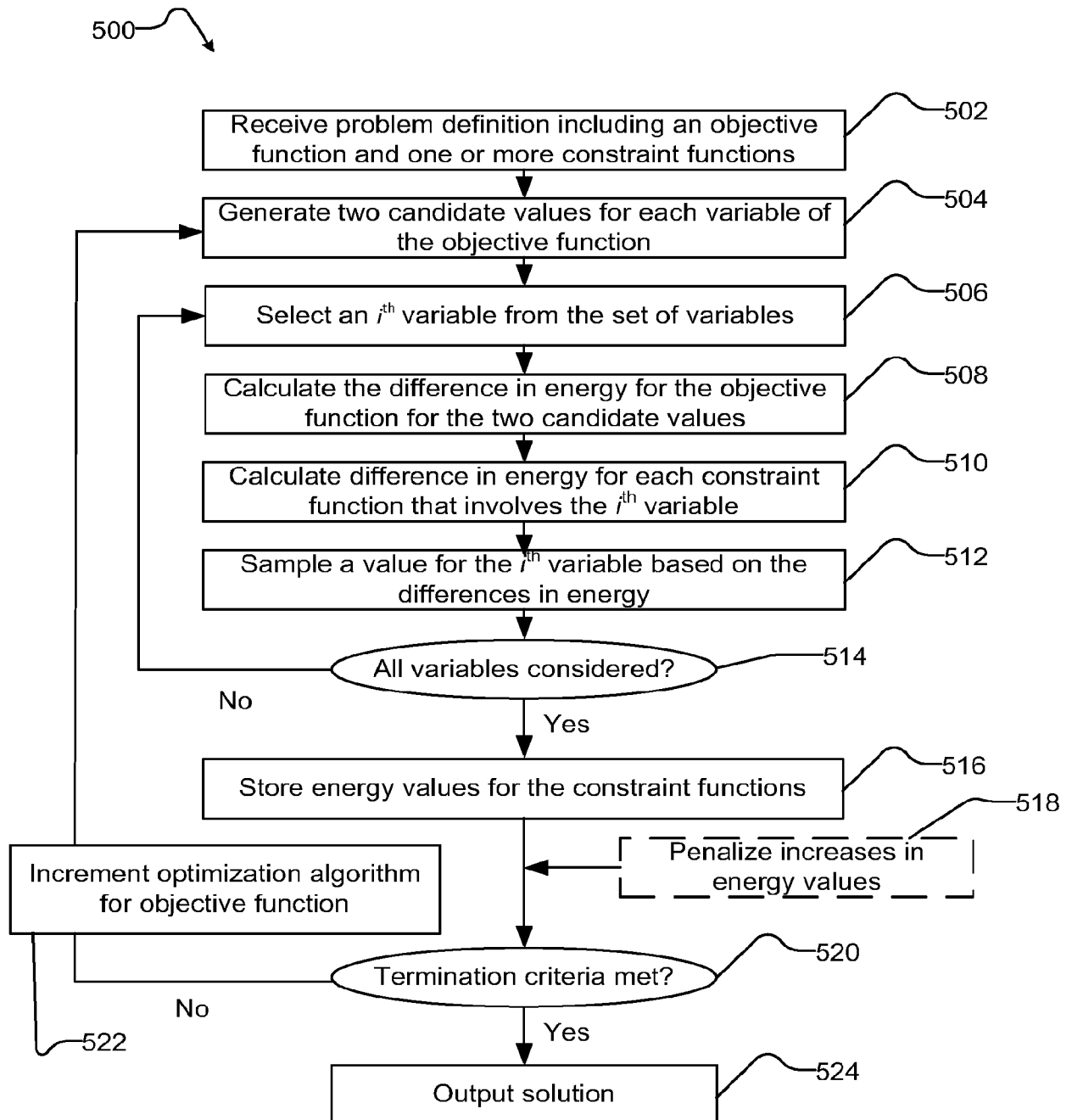


FIGURE 5

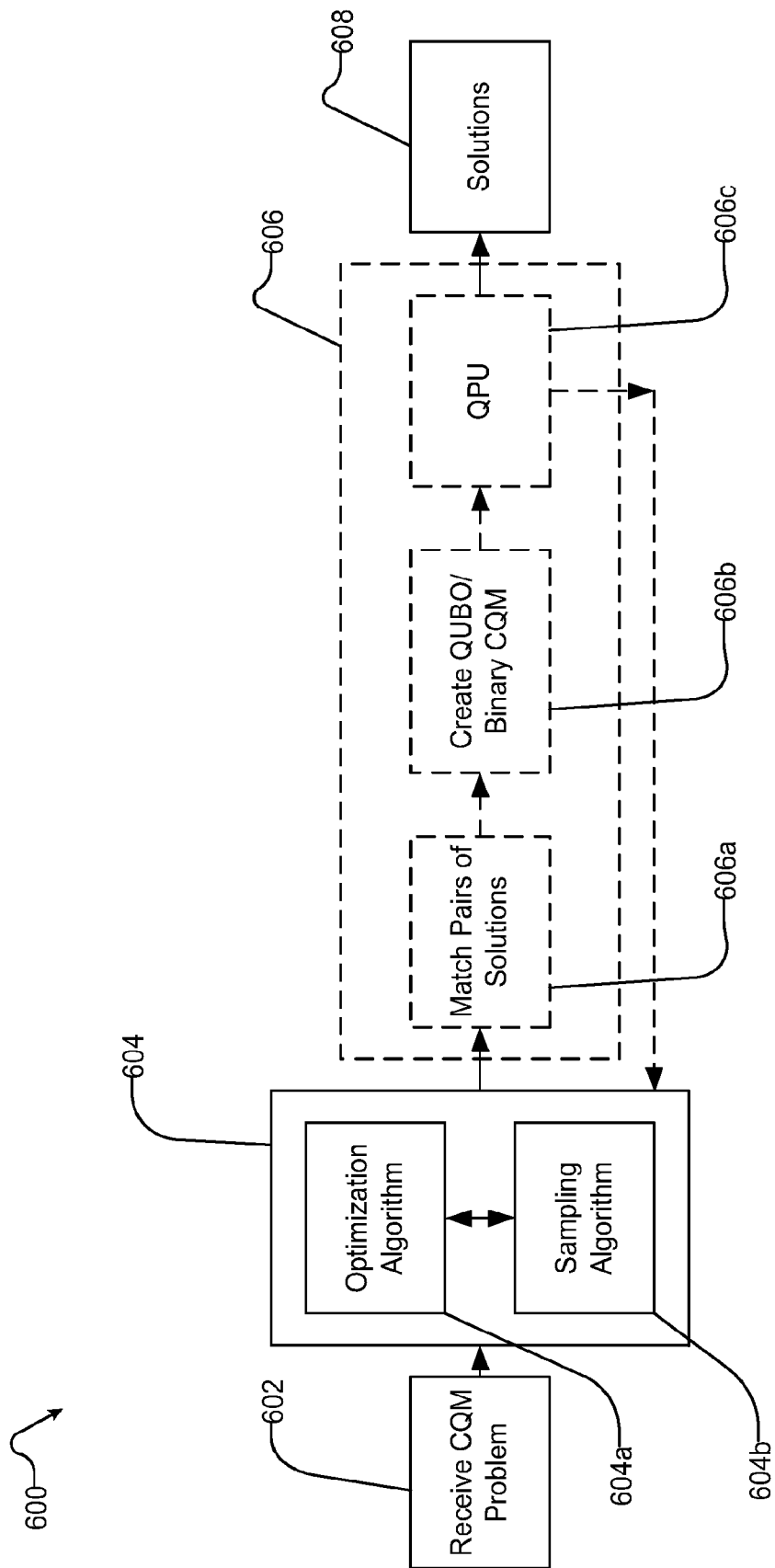


FIGURE 6

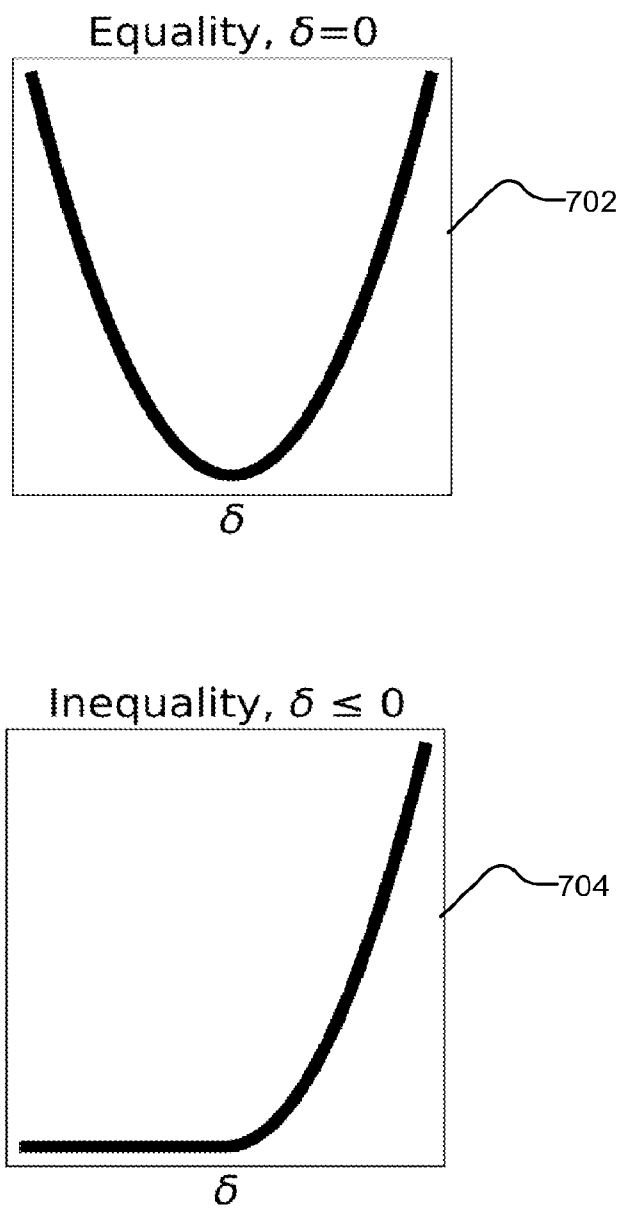


FIGURE 7

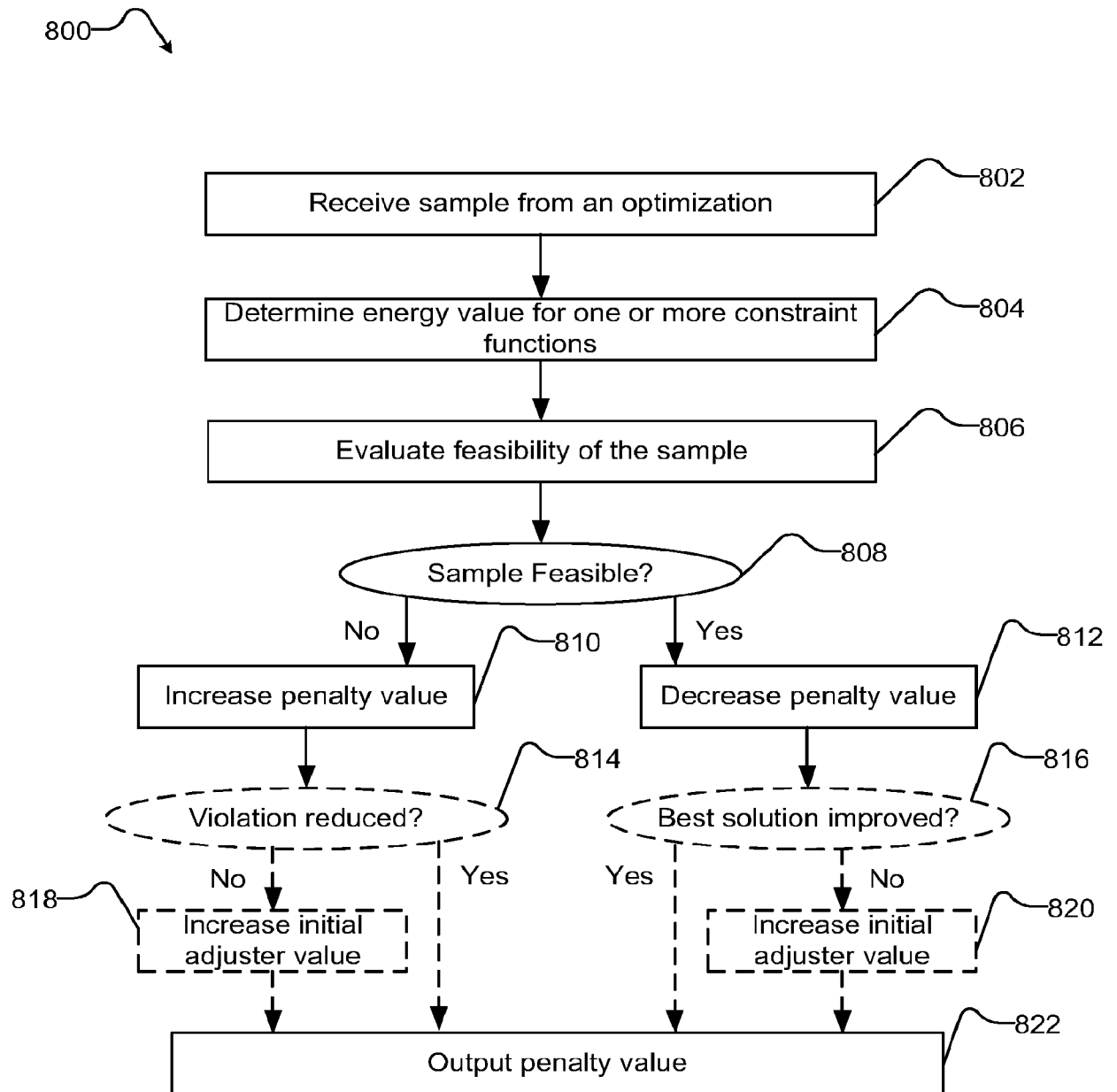


FIGURE 8

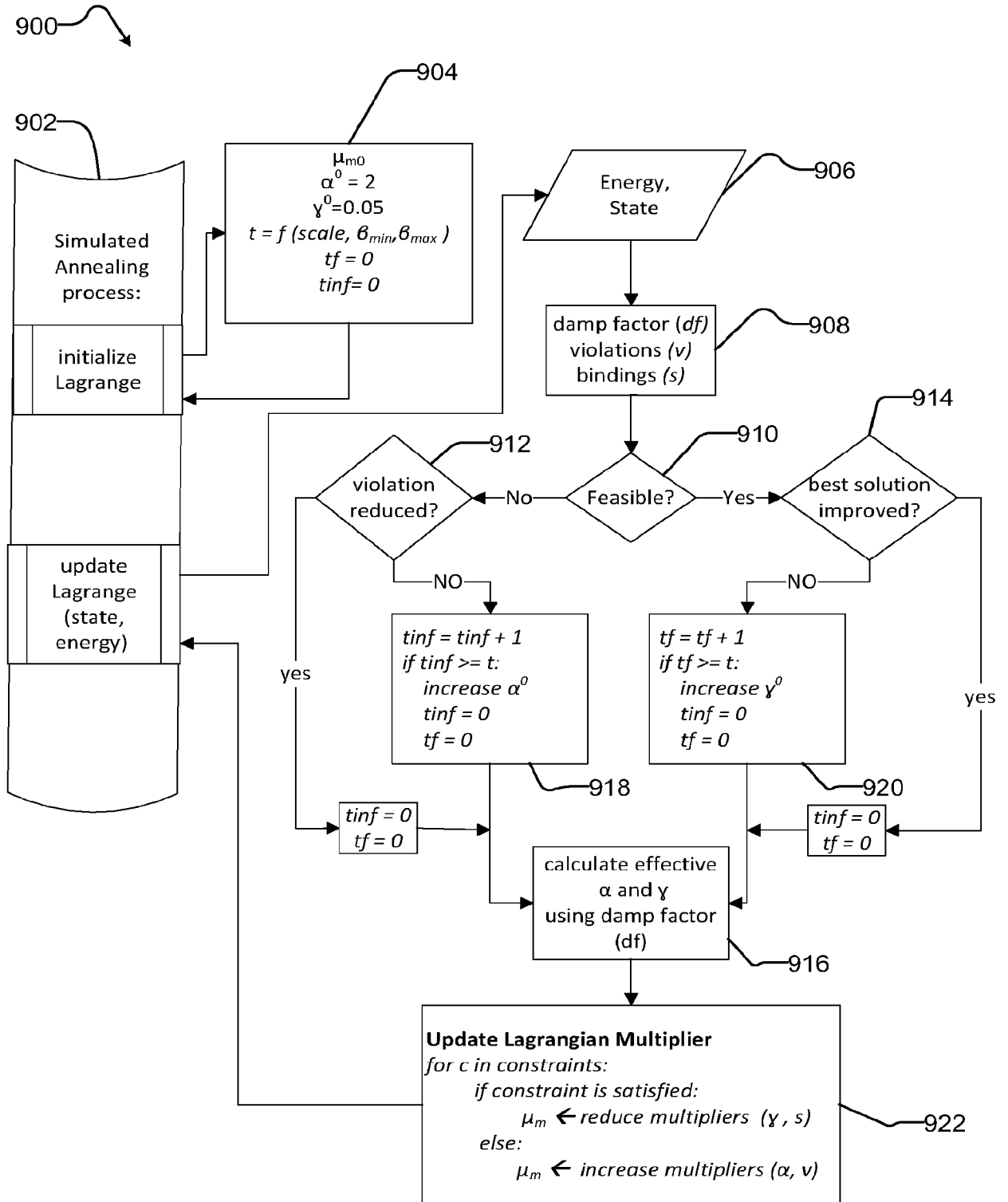


FIGURE 9

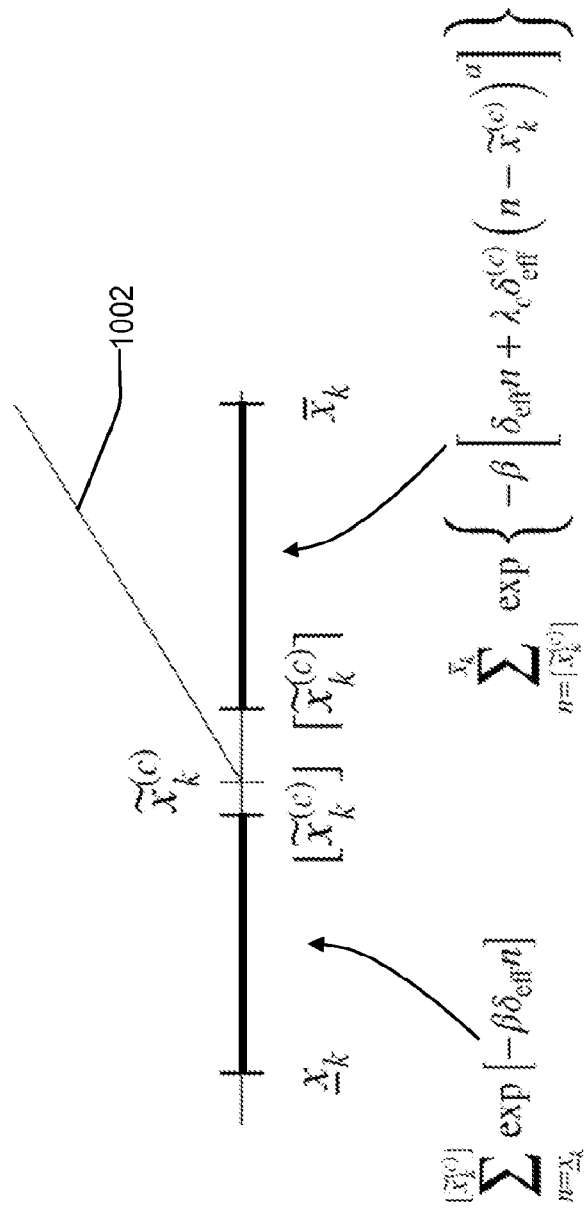


FIGURE 10

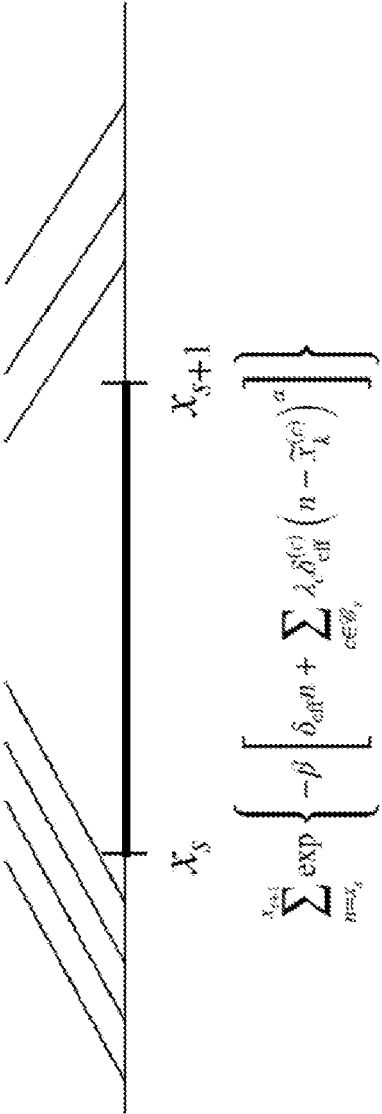


FIGURE 11

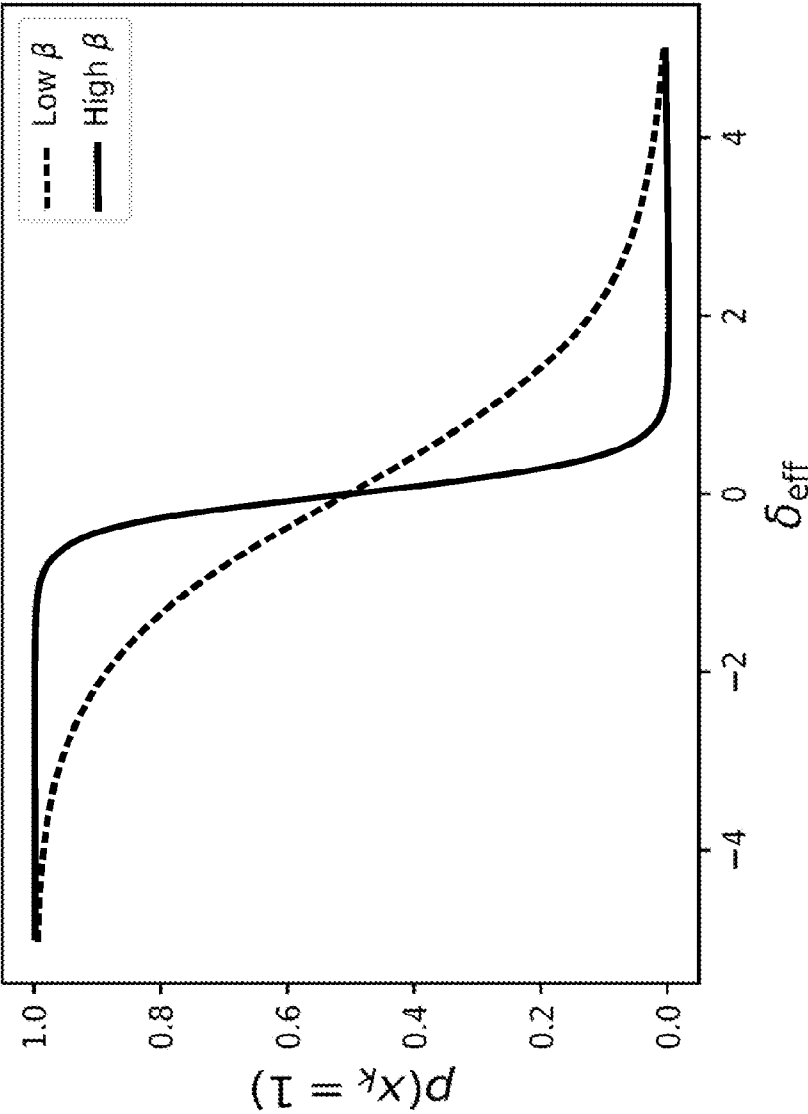


FIGURE 12

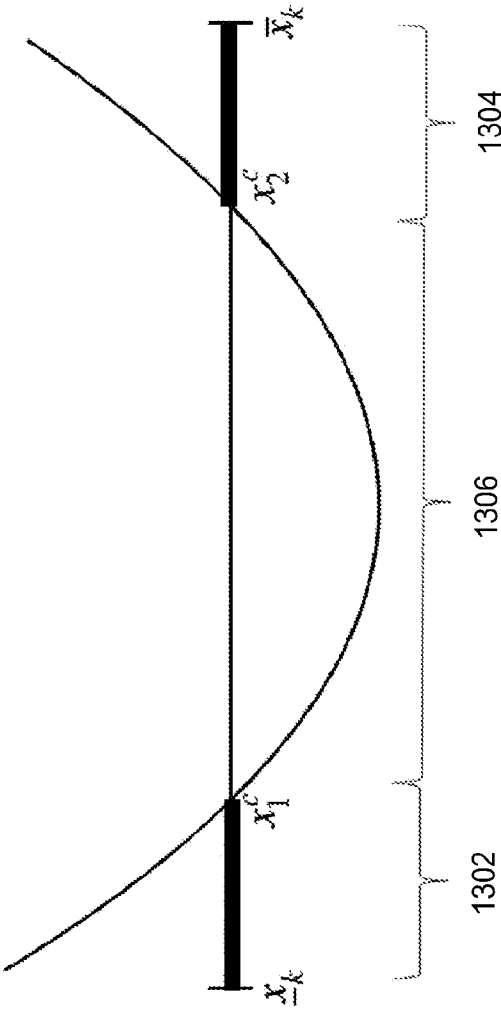


FIGURE 13

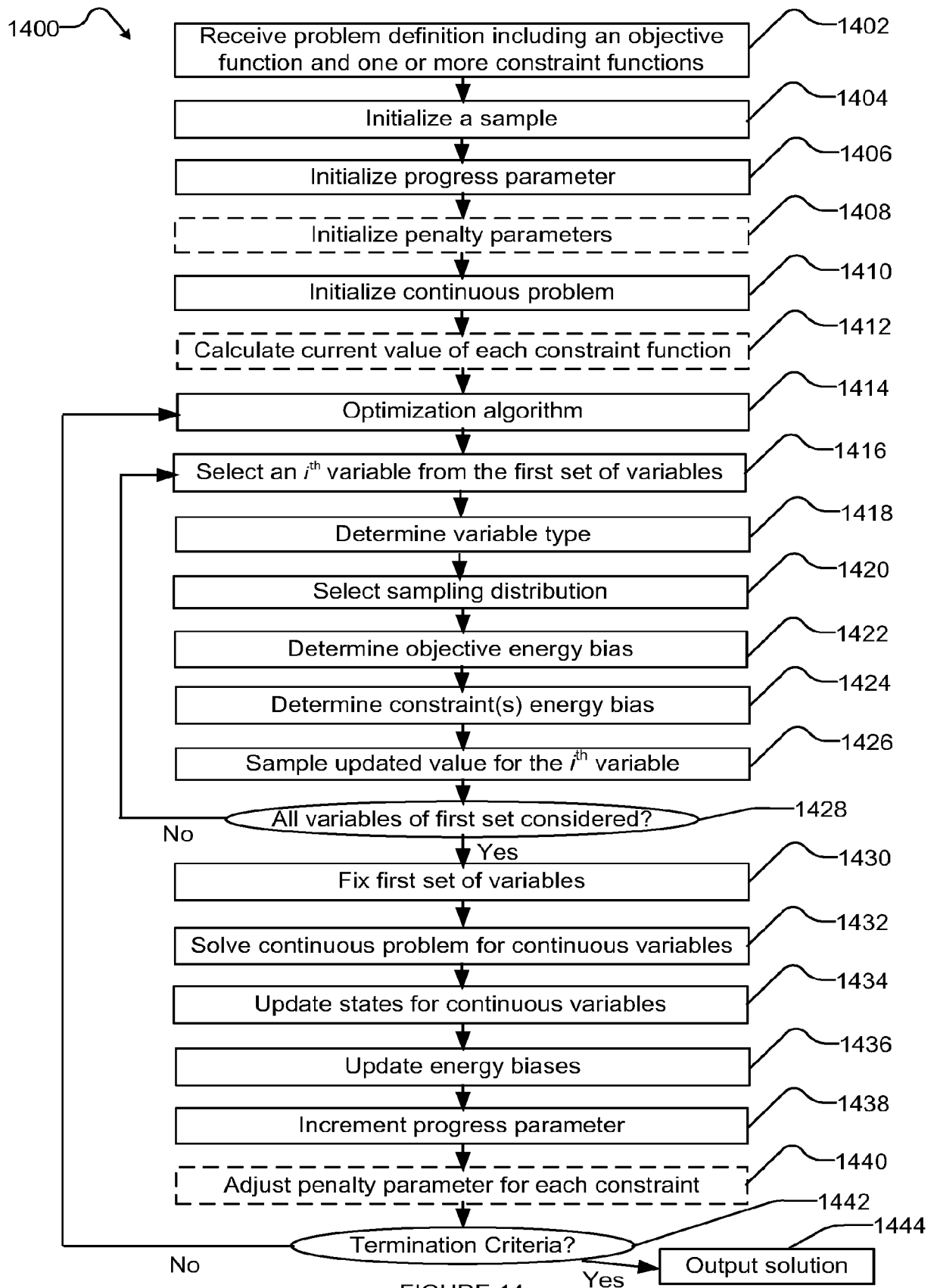


FIGURE 14

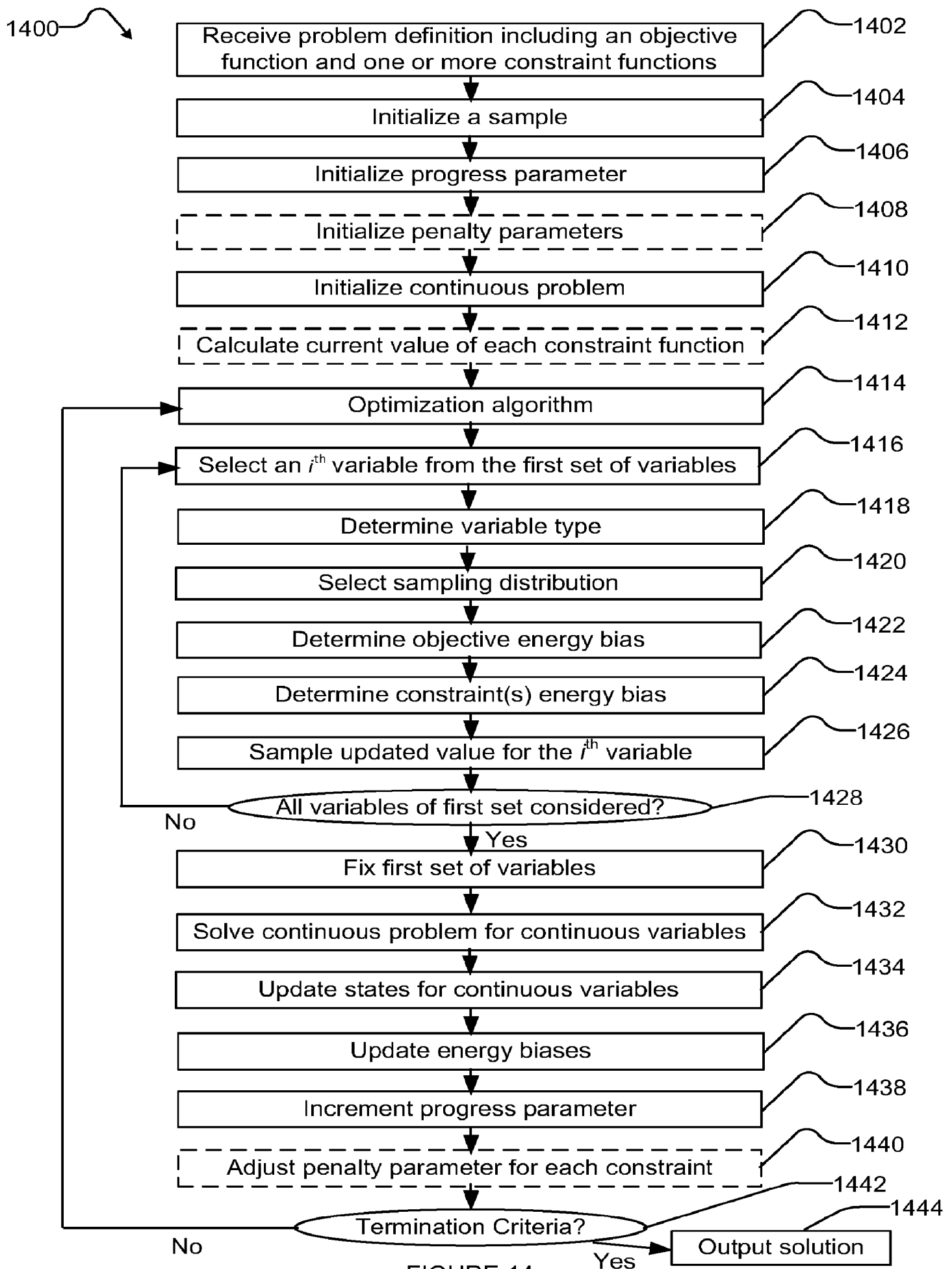


FIGURE 14