Discrete optimization: A quantum revolution?

Stefan Creemers*
IESEG School of Management, Lille, France, s.creemers@ieseg.fr
KU Leuven, Leuven, Belgium, stefan.creemers@kuleuven.be

Luis Fernando Pérez
IESEG School of Management, Lille, France, l.perezarmas@ieseg.fr

We develop several quantum procedures and investigate their potential to solve discrete optimization problems. First, we introduce a binary search procedure and illustrate how it can be used to effectively solve the binary knapsack problem. Next, we introduce two other procedures: a hybrid branch-and-bound procedure that allows to exploit the structure of the problem and a random-ascent procedure that can be used to solve problems that have no clear structure and/or are difficult to solve using traditional methods. We explain how to assess the performance of these procedures and perform a computational experiment. Our results show that we can match the performance of the best classical algorithms when solving the binary knapsack problem. After improving and generalizing our procedures, we show that they can solve any discrete optimization problem using at most $O(\mu\sqrt{2^{nb}})$ operations, where μ is the number of operations required to evaluate the feasibility of a solution, n is the number of decision variables, and 2^b is the number of discrete values that can be assigned to each decision variable. In addition, we demonstrate that our procedures can also be used as heuristics to find (near-) optimal solutions using far less than $O(\mu\sqrt{2^{nb}})$ operations. Not only does our work provide the tools required to explore a myriad of future research directions, it also has the potential to revolutionize the field of discrete optimization.

Keywords: Quantum; computing; algorithm; knapsack; Grover.

1. Introduction

The idea of quantum computing was first launched in 1980, when Benioff defines a quantum mechanical model of a Turing machine. A few years later, Feynman (1982) states the idea of a universal quantum simulator. Building on the work of Benioff and Feynman, Deutsch (1985) describes the first universal quantum computer that is able to efficiently simulate any other quantum computer. In addition, Deutsch also formulates the Deutsch algorithm, the first quantum algorithm that has a proven speedup when compared to classical algorithms. Later, Deutsch and Josza (1992) extend this algorithm. Deutsch and Josza, however, consider a problem (the Deutsch-Josza problem) that has no practical use, and it takes until 1994 for quantum computing to really take off. In 1994, Shor presents a quantum algorithm to find the prime factors of large integers in polynomial time (whereas the best classical algorithm requires sub-exponential time). In theory, Shor's algorithm can be used to break many of the cryptography schemes in use today. Not surprisingly,

*Corresponding author

the publication of Shor's algorithm sparked an enormous interest in quantum computing. A few years later, Grover (1996) presents a quantum algorithm that achieves a quadratic speedup when performing an unstructured search. Arguably, Grover's algorithm is the most important algorithm in quantum computing today. It can be used to solve a multitude of problems and is the corner stone of many other quantum algorithms (see, for example, the Quantum Algorithm Zoo[1] that currently lists 64 quantum algorithms, many of which rely on Grover's algorithm). In what follows, we also focus on Grover-based quantum algorithms, and use them to solve discrete optimization problems.

In this paper, we consider discrete optimization problems that assign discrete values to a set of n decision variables in order to optimize some objective value V. Discrete optimization problems can be formulated as follows:

$$\begin{aligned} &\text{Maximize (or minimize)} \quad V \\ &\text{s.t.} \quad \mathbf{x} \in \Omega \end{aligned}$$

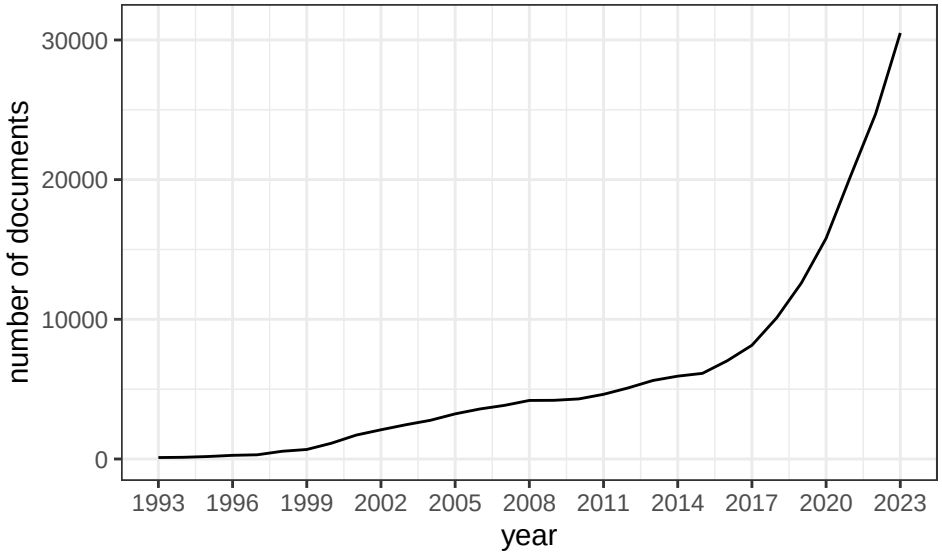
where $\mathbf{x} = \{x_1, \ldots, x_n\}$ is a solution that assigns a discrete value x_i to each decision variable $i : 1 \leq i \leq n$, and Ω is the set of all feasible solutions that respect the logical constraints that define the optimization problem. Whereas it is often easy to evaluate the value and feasibility of a solution $\mathbf{x}$, it can be hard to determine an optimal solution $\mathbf{x}^*$ and its solution value V^*. Discrete optimization problems often have an underlying combinatorial structure, hence the link with combinatorial optimization (Parker and Rardin, 1988). Next to combinatorial optimization problems, integer programming problems and constraint programming problems can also be seen as discrete optimization problems (refer to, for example, Wolsey (2021) and Campbell et al. (2019)).

Discrete optimization problems are at the core of Operations Research (OR), however, they are also mentioned as one of the four use cases of quantum computing by McKinsey (2021). According to McKinsey, the quadratic speedup offered by Grover's algorithm has potential applications in almost every industry. In addition, McKinsey (2023) estimates that quantum computing has an economic potential of up to $1,270 billion (to be realized by 2035 across four industries: chemicals, life sciences, finance, and automotive). They also show that $5.4 billion has already been invested in quantum computing, with companies such as IBM, Google, Amazon, and Microsoft already offering commercial quantum-computing services. Next to the private sector, governments are also investing heavily in quantum technologies: McKinsey (2023) reports that governments have announced up to $34 billion of investments in quantum technologies. In addition, investments in quantum computing are expected to further increase in the upcoming years. This increasing trend can also be seen in the number of published papers that deal with quantum computing. For instance, Figure 1 shows the number of English documents that have been published since 1993 according to Google Scholar when conducting a search using the keyword "quantum computing".

Notwithstanding the tremendous interest in quantum computing, the topic has only received limited attention from the OR community (with only few publications in OR journals that mention "quantum computing"). This is somewhat surprising given the potential that has been ascribed to quantum computers to solve

[1] https://quantumalgorithmzoo.org/

Figure 1. Number of English documents that have been published according to Google Scholar when conducting a search for keyword "quantum computing".



complex optimization problems. It is our goal to (further) open up the field of quantum computing to the OR community, and to show its potential to solve discrete optimization problems. For this purpose, we introduce several new procedures, and demonstrate how these procedures can be used to effectively solve the binary knapsack problem. We also explain how to assess the performance of quantum algorithms and show that our procedures can match the performance of the best classical algorithms when solving the binary knapsack problem. Our procedures, however, can easily be adapted to solve any other discrete optimization problem. Verifying whether we can outperform the best classical algorithms when solving other discrete optimization problems is just one of the myriad of directions for future research that are made possible by our work (refer to the supplementary material for a list of future research topics).

It is beyond the scope of this paper to provide an exhaustive overview of the literature on quantum computing (refer to the Quantum Algorithm Zoo, Nielsen and Chuang (2010), and Hidary (2021) for an overview of the most important results). In addition, even though we briefly touch upon the basics of quantum computing, it is not our goal to provide a comprehensive introduction/tutorial (refer to, for example, Aaronson (2018) for an introductory course and to Nannicini (2020) for a tutorial that does not involve quantum physics). Moreover, in this paper, we do not consider adiabatic quantum computation (see, for example, Albash and Lidar (2018) and Glover et al. (2022) for an introduction). As a result, we do not cover variational quantum algorithms such as the Quantum Approximate Optimization Algorithm (QAOA; see Farhi et al. (2014) for their seminal paper on QAOA, Kurowski et al. (2023) for an example paper that uses QAOA for solving job shop scheduling problems, and Blekos et al. (2023) for a recent review on QAOA and its variants). Instead, we focus on Grover-based algorithms, and provide an OR researcher with the most essential tools that allow to apply quantum computing for optimization purposes. Using these tools, we demonstrate the potential of quantum computing to revolutionize the field of discrete optimization. Note, however, that it may take many more years for this revolution to take place as our procedures require access to a fault-tolerant universal quan-

tum computer that has sufficient scale. Even in the most ambitious roadmap, such a computer is not expected to hit the market before 2030 (refer to, for example, Preskill (2018), MacQuarrie et al. (2020), Babbush et al. (2021), and Cheng et al. (2023) for a discussion).

The remainder of this paper is structured as follows. In Section 2, we introduce the fundamentals of quantum computing as well as the notation that will be used throughout the rest of the paper. Next, in Section 3, we introduce the Deutsch algorithm and its extension to the Deutsch-Josza algorithm. The Deutsch-Josza algorithm serves as an important stepping stone towards Grover's algorithm, that is discussed in Section 4. In Section 5, we explain how Grover's algorithm can be used as a subroutine in a binary search procedure (referred to as `BSP`). This binary search procedure is used in Section 6 to solve 108000 instances of the binary knapsack problem. Because `BSP` is unable to match the performance of the best classical algorithms when solving the binary knapsack problem, we introduce two new procedures: a hybrid branch-and-bound procedure (referred to as `HBB`) in Section 7 and a random-ascent procedure (referred to as `RAP`) in Section 8. We show that these procedures can match the performance of the best classical algorithms that are dedicated to solving the binary knapsack problem. Note, however, that our procedures can easily be adapted to solve any other discrete optimization problem. In Section 9 we demonstrate how `HBB` and `RAP` can further be improved. Section 10 discusses our results, provides ample directions for future research, and concludes.

2. Fundamentals of Quantum Computing

In 1939, Dirac introduced the "bra-ket" notation to facilitate the study of quantum systems. In this notation, a "ket" $|c\rangle$ denotes a column vector c defined in a complex vector space $\mathcal{V}$, a "bra" $\langle r|$ denotes a row vector r that acts as a linear map on $\mathcal{V}$, and a "bra-ket" $\langle r|c\rangle$ represents the inner product of $\langle r|$ and $|c\rangle$. In what follows, we adopt bra-ket notation.

In quantum computing, information is stored in so-called "qubits". A qubit can be seen as a two-state quantum system that has basis states $|0\rangle = (1,0)^\mathsf{T}$ and $|1\rangle = (0,1)^\mathsf{T}$. These basis states are the quantum equivalent of states 0 and 1 of a classical bit. In contrast to a classical bit, however, a qubit may also be in a superposition of both basis states (i.e., it is in both states at the same time). The state of a quantum system is described by a so-called "wave function". In case of a single qubit, the wave function is a linear combination of basis states $|0\rangle$ and $|1\rangle$:

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle\,, \tag{1}$$

where α and β are complex numbers that are also known as the "probability amplitudes" of wave function $|\psi\rangle$. By observing (i.e., measuring) a qubit in superposition, the state of the quantum system collapses into one of the basis states. The probability that a qubit collapses into basis state $|0\rangle$ or $|1\rangle$ is proportional to the square of probability amplitudes α and β, respectively. This result is also known as the Born rule (1926). Given that α and β are complex numbers, we have:

$$|\alpha|^2 + |\beta|^2 = 1. \tag{2}$$

In systems with two (or more) qubits, the basis states can be obtained as the Kronecker product of the

basis states of each individual qubit. For instance, if we consider two qubits, a first basis state is given by:

$$|00\rangle = |0\rangle \otimes |0\rangle = (1,0)^{\mathsf{T}} \otimes (1,0)^{\mathsf{T}} = (1,0,0,0)^{\mathsf{T}}.$$

The other basis states are $|01\rangle = (0,1,0,0)^{\mathsf{T}}$, $|10\rangle = (0,0,1,0)^{\mathsf{T}}$, and $|11\rangle = (0,0,0,1)^{\mathsf{T}}$. If both qubits are in superposition, we get:

$$|\boldsymbol{\Psi}\rangle = |\psi_1\rangle \otimes |\psi_2\rangle = (\alpha_1 |0\rangle + \beta_1 |1\rangle) \otimes (\alpha_2 |0\rangle + \beta_2 |1\rangle) = \alpha_1\alpha_2 |00\rangle + \alpha_1\beta_2 |01\rangle + \beta_1\alpha_2 |10\rangle + \beta_1\beta_2 |11\rangle,$$

where $|\boldsymbol{\Psi}\rangle$ is used to denote the state of a set of qubits.

Similar to classical computing, quantum computing uses logic gates to perform operations. In contrast to the gates of classical computing, however, quantum gates have to be reversible, and are represented by unitary matrices (i.e., a matrix U is unitary if $UU^{-1} = UU^{\mathsf{T}} = I$, where I is the identity matrix). Even though there are many quantum gates, we need to introduce one gate in particular: the Hadmard gate (refer to, for example, Aaronson (2018) for an overview of other quantum gates). The Hadamard gate is used to put qubits in superposition. In case of a single qubit, the Hadamard gate (represented by the letter "H") operates as follows:

$$\mathrm{H}|0\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}\begin{pmatrix}1\\ 0\end{pmatrix} = \frac{|0\rangle + |1\rangle}{\sqrt{2}} = |+\rangle, \quad \mathrm{H}|1\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}\begin{pmatrix}0\\ 1\end{pmatrix} = \frac{|0\rangle - |1\rangle}{\sqrt{2}} = |-\rangle.$$

Because H is an involutory matrix (i.e., a square matrix that is its own inverse), we have $\mathrm{H}|+\rangle = \mathrm{HH}|0\rangle = |0\rangle$ and $\mathrm{H}|-\rangle = \mathrm{HH}|1\rangle = |1\rangle$. Note that a qubit in state $|+\rangle$ or $|-\rangle$ has an equal probability to collapse into basis state $|0\rangle$ or $|1\rangle$ when measured. In general, when applied to a set of n qubits, the Hadamard gate is defined as the n^{th} tensor power of H (denoted $\mathrm{H}^{\otimes n}$), and the probability to measure any of the 2^n basis states equals 2^{-n}.

3. The Deutsch algorithm

The Deutsch algorithm was first published in 1985, has been generalized by Deutsch and Josza in 1992, and has further been improved by Cleve et al. (1998). Even though the Deutsch algorithm solves a problem that has no practical use, it is of particular interest to us because (1) it is the first quantum algorithms for which it has been shown that it can solve a problem more efficiently than a classical algorithm, and (2) it is a stepping stone towards Grover's algorithm that is introduced in the next section.

The Deutsch algorithm solves the following problem: we are given a function f_x that maps a binary input x to a binary output in a way that is unknown to us. Even though function f_x acts as a black box, we are told that f_x is either "constant" or "balanced"; f_x is said to be constant if it always produces the same output, and f_x is balanced if it produces a zero for half of the inputs and a 1 for the other half of the inputs. As such, f_x

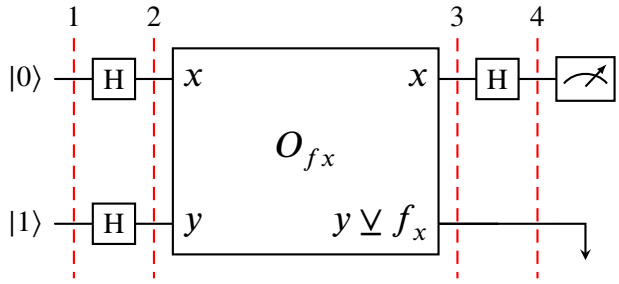
has to be one of the following four functions:

$$
\begin{aligned}
f_0 = 0 \text{ and } f_1 = 0 &\Rightarrow f \text{ is constant (case a)},\\
f_0 = 1 \text{ and } f_1 = 1 &\Rightarrow f \text{ is constant (case b)},\\
f_0 = 0 \text{ and } f_1 = 1 &\Rightarrow f \text{ is balanced (case c)},\\
f_0 = 1 \text{ and } f_1 = 0 &\Rightarrow f \text{ is balanced (case d)},
\end{aligned}
$$

To solve the Deutsch problem, it suffices to determine whether f_x is constant or balanced. The traditional way to solve this problem would be to evaluate both f_0 and f_1, and to observe whether the outputs are the same (i.e., f_x is constant if $f_0 = f_1$, and f_x is balanced otherwise). This solution, however, requires two calls to function f_x. Using the Deutsch algorithm, we only require a single call to function f_x.

The circuit of the Deutsch algorithm is presented in Figure 2 and can be broken down into four steps (indicated by the red dashed lines in the circuit). In a first step, we initialize two qubits, and obtain quantum

Figure 2. Circuit of the Deutsch algorithm.



state $|\Psi_1\rangle = |01\rangle$. Next, in a second step, both qubits are put in superposition using a Hadamard gate:

$$
|\Psi_2\rangle = \mathrm{H}^{\otimes 2}\,|01\rangle = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix} = \frac{1}{2}(|00\rangle - |01\rangle + |10\rangle - |11\rangle) = |+-\rangle\,.
$$

In step 3, we pass both qubits through an oracle gate O_{fx} that, unlike us, has complete knowledge of function f_x. Oracle O_{fx} uses its knowledge of f_x to map state $|xy\rangle$ to $|x(y \veebar f_x)\rangle$, where $\veebar$ represents the XOR operation, and where x and y represent the first and second qubit, respectively. After applying oracle O_{fx}, we get:

$$
|\Psi_3\rangle = \frac{1}{2}\left(|0\rangle \otimes \left(|(0 \veebar f_0)\rangle - |(1 \veebar f_0)\rangle\right) + |1\rangle \otimes \left(|(0 \veebar f_1)\rangle - |(1 \veebar f_1)\rangle\right)\right).
$$

For $x \in \{0, 1\}$, one can verify that $|x\rangle \otimes (|0 \veebar f_x\rangle - |1 \veebar f_x\rangle) = (-1)^{f_x}\,|x\rangle \otimes (|0\rangle - |1\rangle)$, and hence, the sign (or phase) of the second qubit is "kicked back" to the first qubit if $f_x = 1$. As a result, the sign of the

first qubit "flips", and $|\Psi_3\rangle$ may also be written as:

$$|\Psi_3\rangle = \frac{1}{2}\left((-1)^{f_0}|0\rangle \otimes (|0\rangle - |1\rangle) + (-1)^{f_1}|1\rangle \otimes (|0\rangle - |1\rangle)\right).$$

Quantum state $|\Psi_3\rangle$ can further be simplified depending on whether f_x is constant or balanced. For each of the four aforementioned cases, Table 1 lists the mappings of oracle O_{fx}.

Table 1. Mapping of basis states by oracle O_{fx} for each of the four cases.

case	$O_{fx}(\|00\rangle)$	$O_{fx}(\|01\rangle)$	$O_{fx}(\|10\rangle)$	$O_{fx}(\|11\rangle)$
a	$\|00\rangle$	$\|01\rangle$	$\|10\rangle$	$\|11\rangle$
b	$\|01\rangle$	$\|00\rangle$	$\|11\rangle$	$\|10\rangle$
c	$\|00\rangle$	$\|01\rangle$	$\|11\rangle$	$\|10\rangle$
d	$\|01\rangle$	$\|00\rangle$	$\|10\rangle$	$\|11\rangle$

Using these mappings, we get:

$$|\Psi_{3.a}\rangle = \frac{1}{2}(|00\rangle - |01\rangle + |10\rangle - |11\rangle) = \frac{|0\rangle + |1\rangle}{\sqrt{2}} \otimes \frac{|0\rangle - |1\rangle}{\sqrt{2}} = |+-\rangle,$$

$$|\Psi_{3.b}\rangle = \frac{1}{2}(-|00\rangle + |01\rangle - |10\rangle + |11\rangle) = -\frac{|0\rangle + |1\rangle}{\sqrt{2}} \otimes \frac{|0\rangle - |1\rangle}{\sqrt{2}} = -|+-\rangle,$$

$$|\Psi_{3.c}\rangle = \frac{1}{2}(|00\rangle - |01\rangle - |10\rangle + |11\rangle) = \frac{|0\rangle - |1\rangle}{\sqrt{2}} \otimes \frac{|0\rangle - |1\rangle}{\sqrt{2}} = |--\rangle,$$

$$|\Psi_{3.d}\rangle = \frac{1}{2}(-|00\rangle + |01\rangle + |10\rangle - |11\rangle) = -\frac{|0\rangle - |1\rangle}{\sqrt{2}} \otimes \frac{|0\rangle - |1\rangle}{\sqrt{2}} = -|--\rangle.$$

In case a, the state of the system remains unchanged (i.e., $|\Psi_2\rangle = |\Psi_{3.a}\rangle$). In cases b and d, the global phase of the system changes (indicated in blue). The global phase of a system, however, has no impact on the state of the system; $|\Psi\rangle$ and $-|\Psi\rangle$ are said to be equivalent. In cases c and d, the local phase of the first qubit changes (indicated in red). The impact of this change becomes clear in the last step of the Deutsch algorithm. In the last step, we dispose of the second qubit (it is no longer needed) and observe that the first qubit is in state $|+\rangle$ if f_x is constant (cases a and b), and in state $|-\rangle$ if f_x is balanced (cases c and d). Therefore, if the first qubit is passed through a second Hadamard in the last step, we measure $|\Psi_4\rangle = \mathrm{H}|+\rangle = |0\rangle$ if f_x is constant, and $|\Psi_4\rangle = \mathrm{H}|-\rangle = |1\rangle$ if f_x is balanced. As such, the Deutsch algorithm is able to determine whether f_x is constant or balanced with only a single call to oracle O_{fx}. This result was further generalized by Deutsch and Josza, who consider a function $f_{\mathbf{x}}$ that has n binary inputs $\mathbf{x} = \{x_1, \ldots, x_n\}$ (rather than only a single binary input, as was the case in the Deutsch algorithm).

4. Grover's algorithm

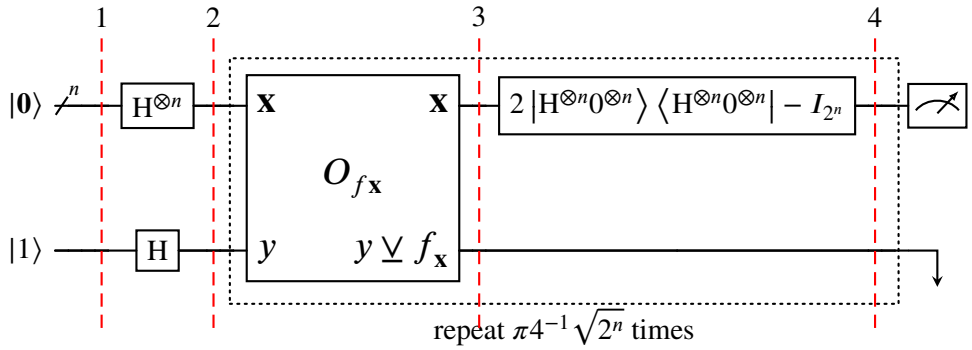
The goal of Grover's algorithm is to find a single target entry in an unstructured database that has 2^n entries. Similar to Deutsch and Josza, Grover considers a function $f_{\mathbf{x}}$ that has n binary inputs that are modeled using n qubits. The resulting quantum system has 2^n basis states that each correspond to an entry of the database. In contrast to Deutsch and Josza, however, $f_{\mathbf{x}}$ is not constant or balanced but returns 1 if the target entry is given as an input, and 0 otherwise. Table 2 provides an example with three binary inputs, resulting in a quantum system that has eight basis states. In this example, function $f_{\mathbf{x}}$ outputs 1 only for target entry 101 (represented by basis state $|101\rangle$). Throughout the remainder of this section, this example will be used to illustrate the different steps of Grover's algorithm.

Table 2. Example database with eight entries that are represented by the basis states of a quantum system that has three qubits. The target entry is printed in blue. Function $f_{\mathbf{x}}$ returns 1 only for the target entry.

$\mathbf{x}$	000	100	010	110	001	101	011	111
$f_{\mathbf{x}}$	0	0	0	0	0	1	0	0

The circuit of Grover's algorithm is presented in Figure 3. Once again, the algorithm can be broken down into four steps (indicated by the red dashed lines in the circuit). The first steps of Grover's algorithm

Figure 3. Circuit of Grover's algorithm.

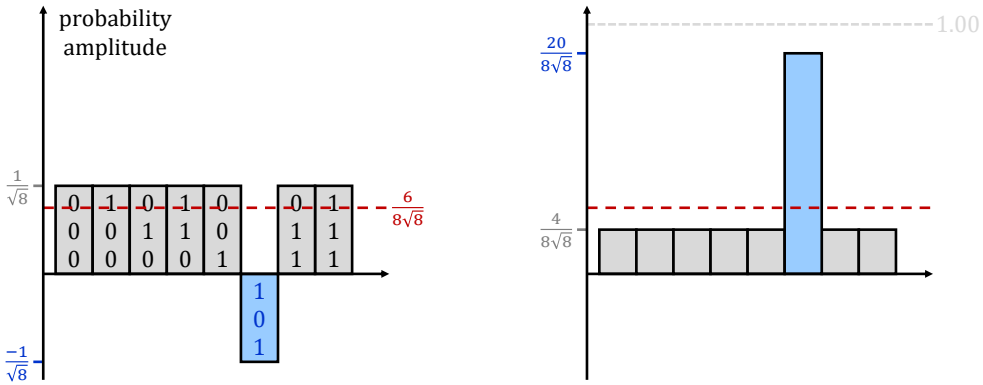


are almost identical to those of the Deutsch-Josza algorithm. The only difference is that oracle $O_{f\mathbf{x}}$ only flips the phase of the basis state that corresponds to the target entry (i.e., oracle $O_{f\mathbf{x}}$ flips the phase of the basis states for which $f_{\mathbf{x}}$ returns 1). In our example, $|\Psi_1\rangle = |0001\rangle$, $|\Psi_2\rangle = |+++-\rangle$, and $|\Psi_3\rangle$ is given by (the basis state whose phase has been flipped is indicated in blue):

$$|\Psi_3\rangle = \frac{1}{\sqrt{8}}\left(|000\rangle + |100\rangle + |010\rangle + |110\rangle + |001\rangle - |101\rangle + |011\rangle + |111\rangle\right) \otimes |-\rangle\,.$$

At this stage, we have only called oracle $O_{f\mathbf{x}}$ once, however, if we measure the first three qubits, there is an equal probability that the system collapses into either one of the 8 basis states. In other words, we have 1 chance out of 8 to find out that $|101\rangle$ is the target basis state; measuring the state of the system at this stage is equivalent to randomly guessing the target entry. To increase the probability of measuring the correct basis

Figure 4. Probability amplitudes of the different basis states before and after applying the diffusion operator in a first iteration (target basis state is colored blue, other basis states are colored gray, and the average amplitude used by the diffusion operator is colored red).



state, Grover uses a so-called "diffusion" operator. In short, the diffusion operator is a quantum gate that reflects the probability amplitudes of the basis states about the average probability amplitude. The goal of the diffusion operator is to amplify the probability amplitude of the target basis state. For three qubits, the diffusion operator is given by:

$$2\left|\mathrm{H}^{\otimes 3}0^{\otimes 3}\right\rangle\left\langle\mathrm{H}^{\otimes 3}0^{\otimes 3}\right| - I_8 = \begin{pmatrix} -0.75 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & -0.75 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & -0.75 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & -0.75 & 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 & -0.75 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & -0.75 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & -0.75 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & -0.75 \end{pmatrix},$$

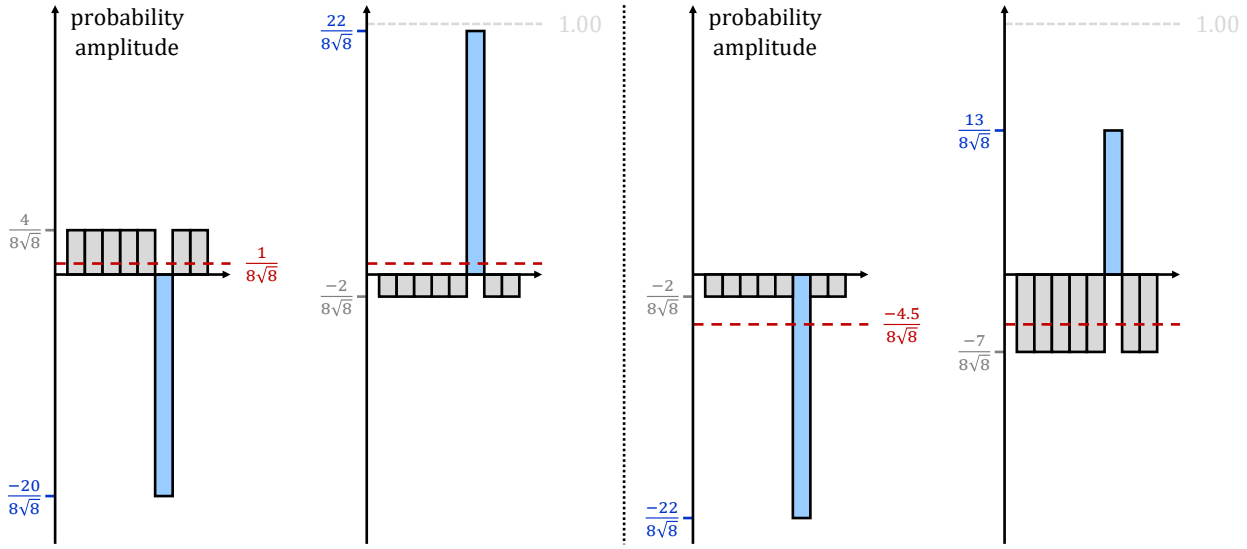
where $\left|0^{\otimes n}\right\rangle$ is the n^{th} tensor power of $|0\rangle$, $|c\rangle\langle r|$ is the outer product of $|c\rangle$ and $\langle r|$, and I_{2^n} is an identity matrix of dimension $2^n \times 2^n$. When applied to our example, one can verify that the average probability amplitude is $6(8\sqrt{8})^{-1}$, and we get:

$$|\Psi_4\rangle = \left(\frac{20}{8\sqrt{8}}|101\rangle + \frac{4}{8\sqrt{8}}(|000\rangle + |100\rangle + |010\rangle + |110\rangle + |001\rangle + |011\rangle + |111\rangle)\right) \otimes |-\rangle\,.$$

In other words, the probability amplitude of the target basis state (i.e., $-(8\sqrt{8})^{-1}$) is reflected about $6(8\sqrt{8})^{-1}$, resulting in probability amplitude $20(8\sqrt{8})^{-1}$. The probability amplitudes of all other basis states (i.e., $(8\sqrt{8})^{-1}$) are also reflected, resulting in probability amplitudes $4(8\sqrt{8})^{-1}$. These reflections are illustrated in Figure 4.

After a first iteration, the probability of identifying the target basis state has increased from 0.125 to

Figure 5. Probability amplitudes of the different basis states before and after applying the diffusion operator in a second (left panel) and third (right panel) iteration (target basis state is colored blue, other basis states are colored gray, and the average amplitude used by the diffusion operator is colored red).



0.7813 (i.e., $(20(8\sqrt{8})^{-1})^2$). We can further increase this probability by repeating steps 3 and 4. We get:

$$|\Psi_{3.2}\rangle = \left(-\frac{20}{8\sqrt{8}}|101\rangle + \frac{4}{8\sqrt{8}}(|000\rangle + |100\rangle + |010\rangle + |110\rangle + |001\rangle + |011\rangle + |111\rangle) \right) \otimes |-\rangle,$$

resulting in an average probability amplitude of $(8\sqrt{8})^{-1}$, and:

$$|\Psi_{4.2}\rangle = \left(\frac{22}{8\sqrt{8}}|101\rangle - \frac{2}{8\sqrt{8}}(|000\rangle + |100\rangle + |010\rangle + |110\rangle + |001\rangle + |011\rangle + |111\rangle) \right) \otimes |-\rangle.$$

After a second iteration, the probability of success increases to 0.9453 (i.e., $(22(8\sqrt{8})^{-1})^2$); we are 94.53 percent certain to identify the target basis state after only two calls to oracle $O_{f\mathbf{x}}$. Another iteration, however, will not increase the success probability:

$$|\Psi_{3.3}\rangle = \left(-\frac{22}{8\sqrt{8}}|101\rangle - \frac{2}{8\sqrt{8}}(|000\rangle + |100\rangle + |010\rangle + |110\rangle + |001\rangle + |011\rangle + |111\rangle) \right) \otimes |-\rangle,$$

resulting in an average probability amplitude of $-4.5(8\sqrt{8})^{-1}$, and:

$$|\Psi_{4.3}\rangle = \left(\frac{13}{8\sqrt{8}}|101\rangle - \frac{7}{8\sqrt{8}}(|000\rangle + |100\rangle + |010\rangle + |110\rangle + |001\rangle + |011\rangle + |111\rangle) \right) \otimes |-\rangle.$$

After a third iteration, the probability of success drops to 0.3301. Figure 5 illustrates the second and third iteration.

Table 3. Average probability amplitude, probability amplitude of the target basis state, probability amplitude of the other basis states, and probability to successfully identify the target basis state before step 4 for different numbers of iterations.

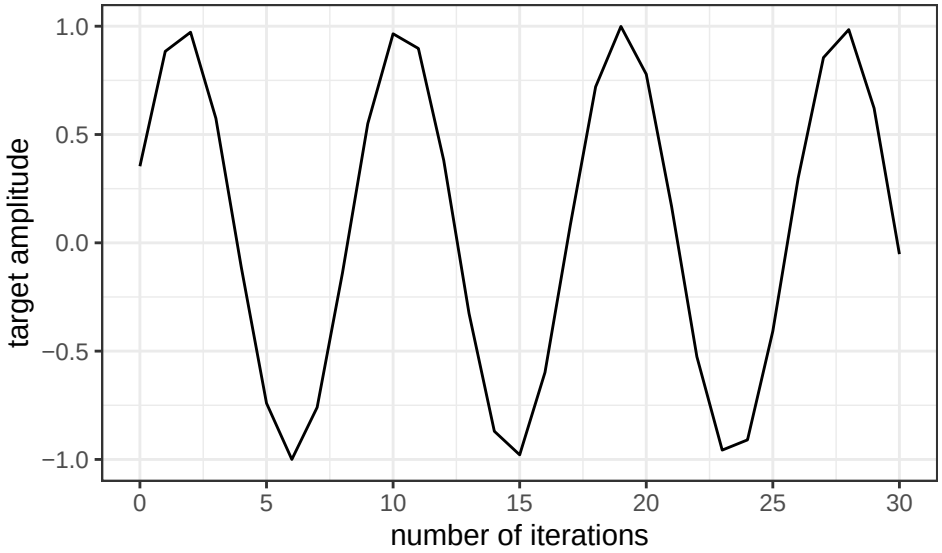
iteration	average amplitude	target amplitude	other amplitude	success probability
1	0.2652	-0.3536	0.3536	0.1250
2	0.0442	-0.8839	0.1768	0.7813
3	-0.1989	-0.9723	-0.0884	0.9453
4	-0.3425	-0.5745	-0.3094	0.3301
5	-0.3149	0.1105	-0.3757	0.0122
6	-0.1298	0.7403	-0.2541	0.5480
7	0.1202	0.9999	-0.0055	0.9998
8	0.3100	0.7596	0.2458	0.5770
9	0.3449	0.1395	0.3743	0.0195
10	0.2073	-0.5504	0.3156	0.3029
11	-0.0339	-0.9650	0.0991	0.9313
12	-0.2582	-0.8972	-0.1669	0.8049
13	-0.3534	-0.3807	-0.3495	0.1450
14	-0.2719	0.3261	-0.3573	0.1063
15	-0.0544	0.8698	-0.1865	0.7566
16	0.1902	0.9787	0.0776	0.9578

For different numbers of iterations, Table 3 lists the average probability amplitude, the probability amplitudes of the target and other basis states, and the probability to successfully identify the target basis state before the diffusion operator is applied. In addition, Figure 6 shows the probability amplitude of the target basis state after each iteration (i.e., after applying the diffusion operator). From Figure 6, one can observe that the probability amplitude of the target basis state can be approximated by a sine function that has its first extremum after two iterations. In fact, to find a target entry in a database of 2^n entries, Grover has shown that the probability amplitude of the target basis state reaches a first extremum after $\pi 4^{-1}\sqrt{2^n}$ iterations (i.e., after $\pi 4^{-1}\sqrt{2^n}$ calls to oracle $O_{f\mathbf{x}}$). A classical algorithm, on the other hand, requires at most 2^n calls to an oracle $O_{f\mathbf{x}}$ to find the target entry in an unstructured database of 2^n entries. As a result, Grover's algorithm achieves a quadratic speedup. In general, if there are m target entries in a database of 2^n entries, Grover's algorithm needs $\pi 4^{-1}\sqrt{m^{-1}2^n}$ calls. Note that, for searching an unstructured database, it is not possible to improve the result of Grover (i.e., a quadratic speedup is the best we can hope to achieve on a quantum computer; see Bennett et al. (1997) for a proof).

5. Using Grover's algorithm to solve the binary knapsack problem

In the previous section, we have seen that Grover's algorithm can be used to perform an unstructured search and can do so more efficiently than a classical algorithm. In this section, we show how to use Grover's algorithm to find the optimal solution of the binary knapsack problem. The binary knapsack problem is an NP-complete problem that has been studied for more than a century (see, for example, Kellerer et al. (2004)). The goal of the binary knapsack problem is to maximize the value of a knapsack by adding items from a pool of n items. Each item $i : 1 \leq i \leq n$ has a weight w_i and a value v_i. The weight of the selected items should not exceed W, the maximum weight capacity of the knapsack. The binary knapsack problem can be

Figure 6. Probability amplitude of the target basis state after each Grover iteration.



formulated as:

$$\text{Maximize} \quad \sum_{i=1}^{n} v_i x_i \tag{3}$$

$$\text{s.t.} \quad \sum_{i=1}^{n} w_i x_i \leq W \tag{4}$$

$$x_i \in \{0, 1\}, \quad \forall\, 1 \leq i \leq n \tag{5}$$

In this formulation, x_i is a binary decision variable that determines whether or not item i has been added to the knapsack, objective function 3 maximizes the value of the knapsack, constraint 4 ensures that the items in the knapsack do not exceed the maximum weight W, and constraint 5 ensures that decision variables are binary. In what follows, we assume $w_i, v_i \in \mathbb{Z}_{\geq 0}$ for all $i : 1 \leq i \leq n$.

To solve the binary knapsack problem, rather than using an omniscient oracle $O_{f_{\mathbf{x}}}$, we have to effectively implement a function $f_{\mathbf{x}}$ that returns 1 if $\mathbf{x}$ is an optimal solution (and 0 otherwise). Next, we can equip Grover's algorithm with $f_{\mathbf{x}}$, and use it flip the phase of all basis states that correspond to an optimal solution (refer to the supplementary material for an implementation of $f_{\mathbf{x}}$ and its use in Grover's algorithm). A solution $\mathbf{x}$ is optimal if it respects the weight constraint and if it has a value of at least V^*:

$$\sum_{i=1}^{n} v_i x_i \geq V^* \tag{6}$$

Unfortunately, however, we don't know V^*. As a result, a procedure is needed to determine V^*. In what follows, we adopt a binary search procedure that evaluates different values V until V^* has been found. For a given value V, we perform $I(n, m) = \left\lceil \pi 4^{-1} \sqrt{m^{-1} 2^n} \right\rceil$ Grover iterations to maximize the probability to

measure one of the m valid solutions (i.e., a solution $\mathbf{x}$ that respects the weight constraint and has a value of at least V). Again, however, we are faced with a problem: we don't know m; we don't know how many solutions have a value of at least V. To resolve this problem, we run Grover's algorithm for different values of $\hat{m}$ until a valid solution is found (where m and $\hat{m}$ denote the real and the assumed number of valid solutions, respectively). To sample different values of $\hat{m}$, we first assume $\hat{m} = 2^n$ (i.e., we assume that all solutions are valid) and evaluate $\hat{m} = \hat{m}/2$ in subsequent runs of Grover's algorithm until a valid solution is found (if no valid solution can be found, we conclude that no solution exists that has a value of at least V and that respects the weight constraint). Algorithm 1 (hereafter referred to as `GUM`) outlines this approach. Even though other approaches have been suggested in the literature (see, for example, Boyer et al. (1996)), `GUM` is of particular interest because: (1) `GUM` requires at most $O(\sqrt{m^{-1}2^n})$ calls to function $f_{\mathbf{x}}$, (2) `GUM` often requires less function calls as a valid solution may be found in early iterations of the procedure, and (3) `GUM` is almost certain to return a valid solution (if it exists; if no valid solution exists, `GUM` returns $\emptyset$). For a more detailed discussion on the performance of `GUM`, refer to the supplementary material.

Algorithm 1: Procedure that uses Grover's algorithm to find a valid solution when the number of valid solutions (m) is unknown in a system that has n binary decision variables.

```
procedure GUM(n)
    m̂ = 2^n;
    do
        perform I(n, m̂) iterations of Grover's algorithm equipped with f_x;
        measure basis state |x⟩;
        if f_x = 1 then
            return x;
        else if m̂ > 1 then
            m̂ = m̂/2;
        else
            return ∅;
    while true;
```

Using `GUM`, we can now introduce the binary search procedure presented in Algorithm 2 (hereafter referred to as `BSP`). After initializing a minimum knapsack value $V_{\min}$ and a maximum knapsack value $V_{\max}$, `BSP` evaluates $L = \lceil \log_2(1 + (V_{\max} - V_{\min})) \rceil$ iterations of an outer loop. In each iteration of the outer loop, we initialize a knapsack value V as the largest integer that is smaller-than-or-equal-to the center of the interval that has endpoints $V_{\min}$ and $V_{\max}$. Next, we run `GUM` to try and measure a valid solution $\mathbf{x}$ for which $f_{\mathbf{x}} = 1$. Procedure `GUM` performs at most $(n+1)$ iterations of an inner loop to evaluate all values of $\hat{m}$. In each of these iterations, we perform $I(n, \hat{m})$ Grover iterations and measure the basis state into which the system collapses (i.e., we measure $|\mathbf{x}\rangle$). Because the outcome of Grover's algorithm is probabilistic, it is possible that solution $\mathbf{x}$ is not valid (e.g., solution $\mathbf{x}$ may have a weight that exceeds W). In addition, if no valid solution exists, any basis state that we measure will correspond to an invalid solution. Therefore, we need to verify whether $\mathbf{x}$ is valid; we need to verify whether $f_{\mathbf{x}} = 1$. If the validity of solution $\mathbf{x}$ has been verified, `GUM` returns a valid solution $\mathbf{x}$ and $V_{\min}$ is updated (i.e., $V_{\min} = V + 1$). If, on the other hand, `GUM` does not

return a valid solution (either because we incorrectly measure an invalid solution and $f_{\mathbf{x}} = 0$, or because no valid solution was found after processing all values of $\widehat{m}$), we update V_{max} (i.e., $V_{\text{max}} = V - 1$). This process continues until $V_{\text{min}} > V_{\text{max}}$. The last found solution and its value are optimal. Note that, in order to reduce the number of iterations, lower- and upper-bound procedures may be used to initialize V_{min} and V_{max}.

Algorithm 2: Binary search procedure (BSP) used to solve the binary knapsack problem using Grover's algorithm.

```
initialize x* = ∅, V* = 0, V_min, V_max, and start procedure BSP(n, V_min, V_max);
procedure BSP(n, V_min, V_max)
    do
        V = ⌊0.5(V_min + V_max)⌋;
        x = GUM(n);
        if x ≠ ∅ then
            update optimal solution x* = x and its value V* = V;
            V_min = V + 1;
            break;
        else
            V_max = V − 1;
            break;
    while V_min ≤ V_max;
```

In what follows, we use an example binary knapsack problem to demonstrate the dynamics of BSP. In the example, we have a pool of three items and a knapsack that has maximum weight capacity $W = 4$. The weights and values of the items are listed in Table 4 (the optimal solution of the example knapsack problem is $\mathbf{x}^* = \{1, 0, 1\}$ and has value $V^* = 5$). The steps of BSP are illustrated in Figure 7. First, we let $V_{\text{min}} = 0$ and $V_{\text{max}} = \sum_{i=1}^{n} v_i = 6$. Next, we initiate a first iteration of the outer loop, and initialize $V = \lfloor 0.5(V_{\text{min}} + V_{\text{max}}) \rfloor = 3$. In a first iteration of the GUM inner loop, we let $\widehat{m} = 2^n = 8$ and perform $I(n, \widehat{m}) = 1$ Grover iteration. Given $V = 3$, there are two valid solutions (i.e., solution $\{1, 0, 0\}$ and $\{1, 0, 1\}$). As such, the phases of basis states $|100\rangle$ and $|101\rangle$ are flipped by Grover's algorithm. One can verify that, after a single Grover iteration, we have a 50 percent probability to measure $|100\rangle$ and a 50 percent probability to measure $|101\rangle$. In other words, if we observe the system after one Grover iteration, we are certain to measure a valid solution that has at least value $V = 3$. After verifying that $\{1, 0, 0\}$ (or $\{1, 0, 1\}$) is a valid solution, we let $V_{\text{min}} = V + 1 = 4$, update the best-found solution, and start a second iteration of the outer loop. In this second iteration, we let $V = 5$. Next, we initiate a first iteration of the inner loop, let $\widehat{m} = 8$ and again perform $I(n, \widehat{m}) = 1$ Grover iteration. Given $V = 5$, however, one can verify that there is only one valid solution (i.e., solution $\{1, 0, 1\}$). If there is only a single valid solution, a single Grover iteration only has a 78.13 percent chance to measure the target basis state (see also Table 3 in the previous section). In other words, there is a 21.88 percent chance that we measure a basis state that is not valid. This illustrates why every solution resulting from Grover's algorithm needs to be verified. If the validity of the solution has been verified, we update the best-found solution, let $V_{\text{min}} = 6$, and start a third iteration of the outer loop in which $V = 6$ (in this third iteration, no valid solution can be found for any value of $\widehat{m}$, and we let $V_{\text{max}} = V - 1 = 5$;

Table 4. Data for example knapsack problem with $n = 3$ items.

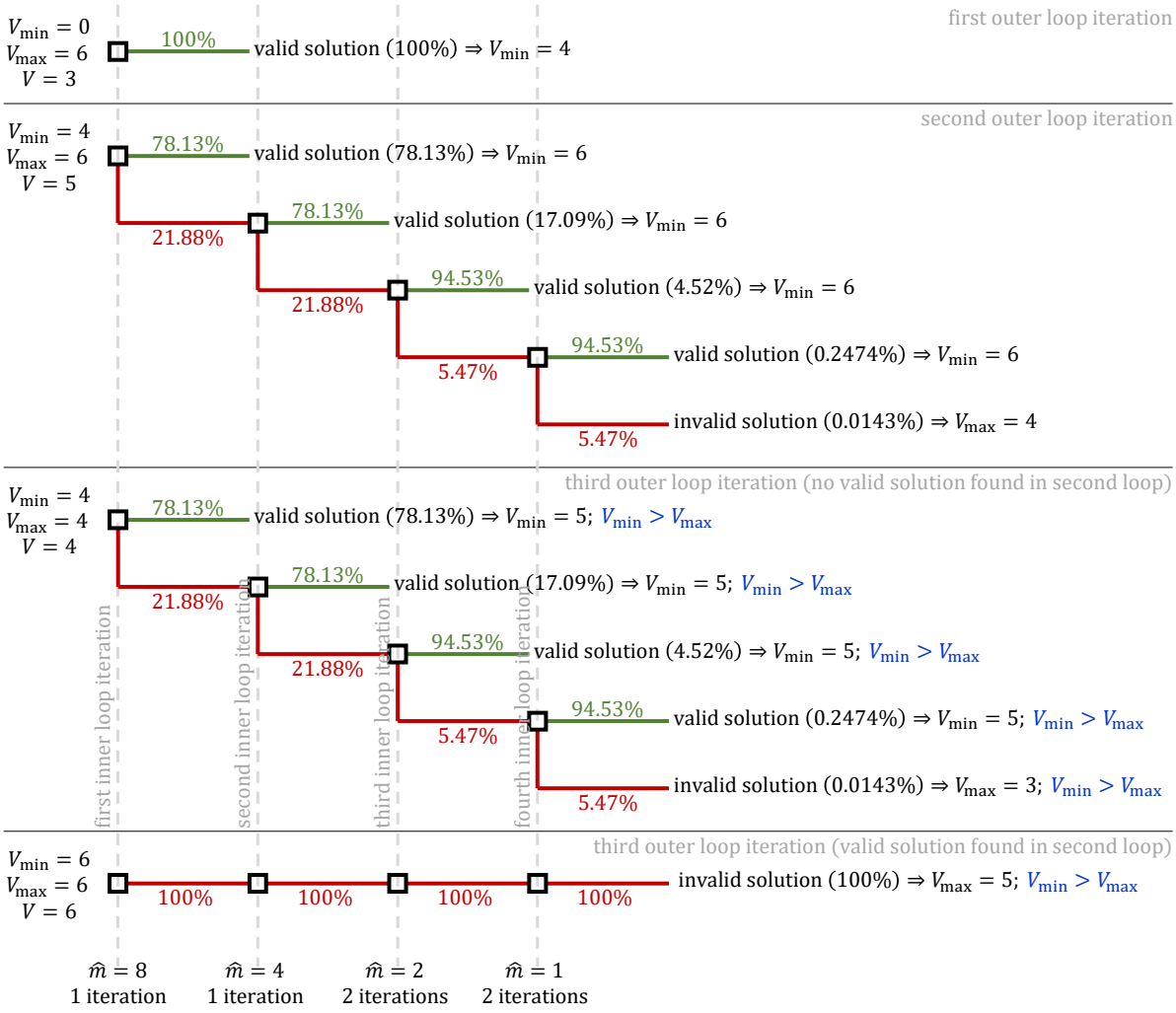
i	w_i	v_i
1	2	3
2	3	1
3	2	2
$W = 4$		

as a result, $V_{\min} > V_{\max}$, and the procedure stops). If, on the other hand, no valid solution was identified (with probability 0.2188), we let $\widehat{m} = \widehat{m}/2 = 4$, and we perform $I(n, \widehat{m}) = 1$ Grover iteration. Again, we have a 78.13 percent chance to measure the basis state that corresponds to valid solution $\{1, 0, 1\}$. As such, we have a 17.09 percent chance to end up a situation where we did not measure a valid solution in the first iteration of the inner loop but did measure a valid solution in the second iteration of the inner loop. If a valid solution was measured, we update the best-found solution, let $V_{\min} = 6$, and start a third iteration of the outer loop in which $V = 6$. If no valid solution was measured, we let $\widehat{m} = 2$, and perform $I(n, \widehat{m}) = 2$ Grover iterations. With 2 iterations, we have a 94.53 percent chance to identify a valid solution. If, however, we still fail to identify a valid solution, we let $\widehat{m} = 1$, and again perform $I(n, \widehat{m}) = 2$ Grover iterations. In total, we are 99.99 percent certain to measure $|101\rangle$, and to obtain $\{1, 0, 1\}$ as a valid solution during the second iteration of the outer loop. There is, however, a 0.0143 percent probability to measure the wrong basis state. If this happens, verification of the solution fails, we cannot reduce $\widehat{m}$ any further, and we let $V_{\max} = V - 1 = 4$ (i.e., we have wrongfully concluded that $V = 5$ has no valid solution; this illustrates why it is possible that we are unable to identify an optimal solution if Grover's algorithm is used in an optimization procedure). At this point, we enter a third iteration of the outer loop with $V = 4$ (here again, we have a 99.99 percent probability to identify $\{1, 0, 1\}$ as the only valid solution that has value $V \geq 4$, after which we update $V_{\min}$ such that $V_{\min} > V_{\max}$, and the procedure stops; conversely, we have a 0.0143 percent probability to identify an invalid solution, after which we update $V_{\max}$ such that $V_{\min} > V_{\max}$, and the procedure also stops). Even though we are almost certain to find the optimal solution, we also require at least 8 Grover iterations (and at most 13). In addition, we need to verify the validity of at least 6 knapsack solutions (and at most 9). For each Grover iteration, we actually require two calls to function $f_{\mathbf{x}}$ (one call to function $f_{\mathbf{x}}$ and one call to reverse function $f_{\mathbf{x}}^{\dagger}$ to "uncompute"; refer to the supplementary material for additional details). Each validation also requires a call to function $f_{\mathbf{x}}$. Therefore, to solve the example knapsack problem using BSP, we need at least 22 calls to function $f_{\mathbf{x}}$, and at most 35 (note that a brute-force classical approach would require us to evaluate only 8 knapsack solutions).

6. Performance of BSP

In the previous section, we have seen that the quadratic speedup of Grover's algorithm doesn't always result in a speedup when solving discrete optimization problems. To further investigate whether BSP can offer a speedup when compared to classical algorithms, we first have a look at the complexity of BSP. To solve a discrete optimization problem that has n binary decision variables and optimal objective value V^*, BSP

Figure 7. Illustration of the dynamics of procedure BSP for the example knapsack problem. In each node (represented by a square), we perform 1 to 2 Grover iterations (depending on the value of $\hat{m}$), we measure the basis state, and we verify whether the corresponding solution is valid. If the solution is valid (indicated in green), $V_{\min}$ is updated. If, on the other hand, the solution is not valid (indicated in red), we update $\hat{m}$ and/or $V_{\max}$ (if $\hat{m} = 1$). The procedure stops if $V_{\min} > V_{\max}$ (indicated in blue).



performs L iterations of an outer loop, where L is a logarithmic function of the bounds on V^*. For each iteration of the outer loop, BSP performs up to $(n+1)$ runs of Grover's algorithm. For each run $i : 0 \leq i \leq n$, we perform $I(n, 2^{(n-i)})$ Grover iterations. Each Grover iteration requires two calls to function $f_{\mathbf{x}}$. In addition, after each run of Grover's algorithm, we need to verify whether the measured solution is valid, resulting in another call to function $f_{\mathbf{x}}$. In total, BSP requires at most $L \sum_{i=0}^{n} \left(1 + 2I(n, 2^{(n-i)})\right) \approx O(L\sqrt{2^n})$ calls to $f_{\mathbf{x}}$. If we consider the binary knapsack problem, function $f_{\mathbf{x}}$ requires $O(n)$ operations. As a result, BSP requires at most $O(Ln\sqrt{2^n})$ operations to solve a binary knapsack problem that has n items. In practice, however, BSP may run much faster because not all $(n+1)$ runs of Grover's algorithm may be required (i.e., a valid solution may be measured in run $i < (n+1)$; see also the discussion on the performance of GUM in the supplementary material).

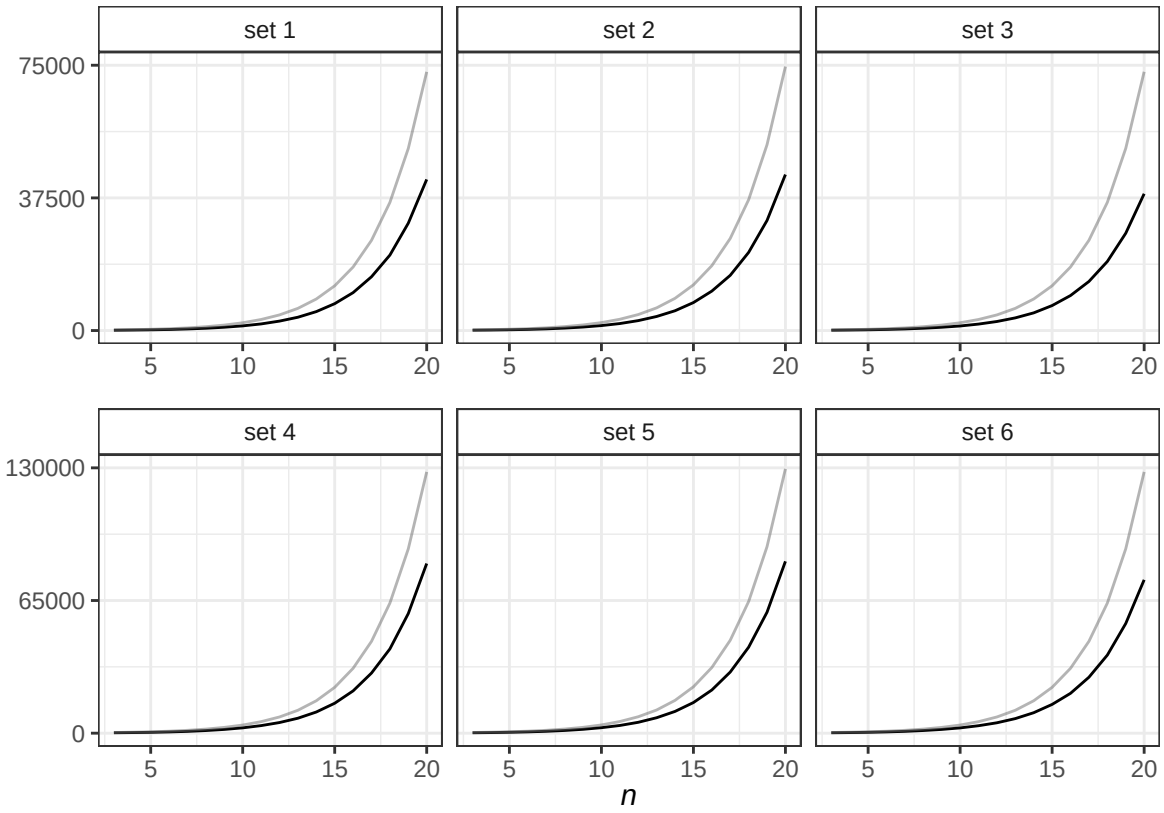
To verify the practical runtime of BSP, we solve a set of 108000 instances of the binary knapsack problem. We use the instance generator of Pisinger (2005) to generate six sets of 1000 problems for each value of $n : n \in \{3, 4, \ldots, 19, 20\}$:

- Set 1 has uncorrelated values and weights that range from 1 to 1000.
- Set 2 has strongly correlated values and weights that range from 1 to 1000.
- Set 3 has weights that range from 1 to 1000 and values $v_i = 3\lceil w_i/3 \rceil$ for each item i (so-called "profit-ceiling instances").
- Set 4 has uncorrelated values and weights that range from 1 to 1000000.
- Set 5 has strongly correlated values and weights that range from 1 to 1000000.
- Set 6 has weights that range from 1 to 1000000 and values $v_i = 3\lceil w_i/3 \rceil$ for each item i.

The "hardness" of each set is determined by the type of instances, and by the range of values and weights. Instances that have uncorrelated values and weights are considered to be very easy to solve. Instances that have strongly correlated values and weights are considered to be hard to solve. The so-called "profit-ceiling" instances are considered very hard to solve. A large range of weights and values also makes instances more difficult to solve.

We use a Discrete-Time Markov Chain (DTMC) to assess the performance of BSP when solving the 108000 instances (see Appendix B for a detailed discussion on how to evaluate the performance of a quantum algorithm). The results of the experiment are summarized in Table 5 and Figure 8. Table 5 reports the minimum probability to find an optimal solution for each problem set and for each value of $n : 3 \leq n \leq 20$. As expected, BSP can be used with confidence to solve all but the smallest binary knapsack problems. Figure 8, on the other hand, shows the expected worst-case number of function calls as well as the expected number of function calls when solving 1000 binary knapsack problems for each problem set and for each value of n. From Figure 8, we can conclude that: (1) the type of problem instance only has limited impact on the performance of the procedure (a discussion on the slight differences among the different problem sets is available in Section 9), (2) the range of the values has a logarithmic effect on the expected number of function calls, and (3) the expected number of function calls is much smaller than the worst-case number of function

Figure 8. Comparison of average worst-case number of function calls (gray) and average number of expected function calls (black) when using procedure BSP for solving 1000 binary knapsack problems for each problem set and for each value of $n : 3 \leq n \leq 20$. Note that instances of sets 4, 5, and 6 (that have values ranging from 1 to 1000000) require twice as many operations to solve as instances of sets 1, 2, and 3 (that have values ranging from 1 to 1000); indicating a logarithmic effect.



calls. To further explore the practical performance of BSP, we return to Table 5 that reports the worst-case number of function calls as well as the expected number of function calls divided by $L\sqrt{2^n}$. It seems that, in the worst case, BSP requires $\approx 5L\sqrt{2^n}$ calls to function $f_{\mathbf{x}}$. In practice, however, only $\approx 3L\sqrt{2^n}$ calls are required. Even though this is a significant reduction in the number of function calls, this implies that, in practice, BSP still needs $O(nL\sqrt{2^n})$ operations when solving binary knapsack problems. The best classical algorithms, on the other hand, require $O(n2^b)$ or $O(n\sqrt{2^n})$ operations (where b is the number of bits required to represent W; see also Kellerer et al., 2004). As such, BSP is unable to match the performance of the best classical algorithms. To keep things simple, in what follows, we consider that the best classical algorithms require $O(n\sqrt{2^n})$ operations (i.e., we assume that W is sufficiently large such that $n2^b > n\sqrt{2^n}$).

7. Hybrid branch-and-bound procedure

In this section, we introduce a new hybrid branch-and-bound procedure (hereafter referred to as HBB) that, in contrast to BSP, is able to match the performance of the best classical algorithms for solving the binary knapsack problem. In what follows, we use HBB to solve the binary knapsack problem. An outline of HBB is

Table 5. Minimum probability to find an optimal solution, average worst-case, and average expected number of function calls divided by $\sqrt{2^n}$ when using procedure BSP for solving 1000 binary knapsack problems for each problem set and for each value of n : $3 \leq n \leq 20$.

n	minimum probability						$E\left[WCcalls/L\sqrt{2^n}\right]$						$E\left[calls/L\sqrt{2^n}\right]$					
	set 1	set 2	set 3	set 4	set 5	set 6	set 1	set 2	set 3	set 4	set 5	set 6	set 1	set 2	set 3	set 4	set 5	set 6
3	0.91	0.93	0.93	0.93	0.93	0.95	5.41	5.40	5.41	5.53	5.53	5.53	3.13	3.35	3.24	3.33	3.43	3.40
4	0.97	0.99	0.97	0.99	0.99	0.99	5.50	5.50	5.50	5.61	5.61	5.61	3.18	3.34	3.29	3.38	3.46	3.45
5	0.98	0.98	0.98	0.98	0.98	0.98	5.40	5.44	5.41	5.52	5.54	5.52	3.13	3.32	3.27	3.39	3.50	3.47
6	0.99	0.99	0.99	0.99	0.99	0.99	5.40	5.42	5.40	5.51	5.52	5.51	3.21	3.35	3.30	3.48	3.56	3.53
7	0.99	0.99	0.99	0.99	0.99	0.99	5.45	5.42	5.45	5.54	5.53	5.54	3.29	3.38	3.32	3.54	3.61	3.58
8	0.99	1.00	1.00	0.99	0.99	0.99	5.47	5.43	5.47	5.56	5.54	5.56	3.30	3.40	3.30	3.61	3.65	3.62
9	1.00	1.00	1.00	1.00	1.00	1.00	5.41	5.42	5.41	5.52	5.52	5.52	3.28	3.39	3.23	3.60	3.65	3.58
10	1.00	1.00	1.00	1.00	1.00	1.00	5.35	5.40	5.35	5.45	5.48	5.45	3.22	3.37	3.16	3.58	3.65	3.53
11	1.00	1.00	1.00	1.00	1.00	1.00	5.34	5.42	5.34	5.44	5.48	5.44	3.20	3.36	3.12	3.57	3.62	3.52
12	1.00	1.00	1.00	1.00	1.00	1.00	5.32	5.37	5.32	5.40	5.43	5.40	3.22	3.34	3.08	3.54	3.60	3.46
13	1.00	1.00	1.00	1.00	1.00	1.00	5.31	5.30	5.32	5.38	5.39	5.38	3.19	3.29	3.04	3.51	3.59	3.43
14	1.00	1.00	1.00	1.00	1.00	1.00	5.30	5.24	5.30	5.37	5.34	5.37	3.19	3.22	2.97	3.51	3.54	3.39
15	1.00	1.00	1.00	1.00	1.00	1.00	5.27	5.19	5.27	5.34	5.30	5.34	3.16	3.18	2.94	3.49	3.50	3.35
16	1.00	1.00	1.00	1.00	1.00	1.00	5.23	5.16	5.23	5.31	5.27	5.31	3.12	3.14	2.88	3.46	3.46	3.26
17	1.00	1.00	1.00	1.00	1.00	1.00	5.18	5.15	5.18	5.29	5.25	5.29	3.09	3.09	2.82	3.46	3.45	3.22
18	1.00	1.00	1.00	1.00	1.00	1.00	5.15	5.16	5.15	5.26	5.25	5.26	3.03	3.08	2.78	3.41	3.43	3.16
19	1.00	1.00	1.00	1.00	1.00	1.00	5.14	5.18	5.14	5.24	5.26	5.24	3.03	3.07	2.74	3.40	3.41	3.12
20	1.00	1.00	1.00	1.00	1.00	1.00	5.14	5.20	5.14	5.24	5.27	5.24	3.00	3.08	2.71	3.40	3.43	3.07

given by Algorithm 3. Procedure HBB uses a recursive depth-first strategy to search a tree that has n levels. At each level $i : 1 \leq i \leq n$, we verify whether item i can be added to the current (partial) solution $\mathbf{x}_{1(i-1)} = \{x_1, \ldots, x_{(i-1)}\}$ that has value $V = \sum_{j=1}^{(i-1)} x_j v_j$ and remaining weight capacity $R = W - \sum_{j=1}^{(i-1)} x_j w_j$. If the weight of item i does not exceed the remaining weight capacity, we add item i (i.e., we let $x_i = 1$) and let $\mathbf{x}_{1i} = \{x_1, \ldots, x_{(i-1)}\} \cup \{x_i\}$, $V' = V + v_i$, and $R' = R - w_i$. In addition, if $V' > V^*$, we also update the best-found solution and its value. Next, we use GUM to assess whether a better solution can still be found at a lower level of the search tree. If such a solution $\mathbf{x}_{(i+1)n}$ can be found, we update the best-found solution and enter a new recursion in which item i is added. After (potentially) exploring the branch in which we add item i, we also explore the branch in which we do not add item i (i.e., in which $x_i = 0$). At a given level of the tree, ω items remain to be added. For small ω, however, it is possible that GUM (that relies on Grover's algorithm) is unable to measure a valid solution (even if that solution exists; refer to the supplementary material for a discussion on the limitations of Grover's algorithm). In addition, for small ω, the quadratic speedup (and hence the computational gain) obtained by Grover's algorithm is limited. Therefore, we propose a "hybrid" procedure that uses: (1) GUM if more than ω items remain, and (2) a classical approach if ω items remain. If ω items remain, the classical approach determines the set of remaining items in Ω_{in} (i.e., the set of all partial solutions that consider items $j \geq i$) that best improves the current (partial) solution without violating the maximum weight constraint. If such a solution ($\mathbf{x}^\star_{in}$; with value $V^\star_{in}$ and weight $W^\star_{in}$) can be found, we update the best-found solution. A hybrid approach has already been proposed by, for example, Ambainis et al. (2018) who describe an algorithm that is dedicated to solving the travelling salesman problem. Procedure HBB is also similar to the backtracking algorithm of Montanaro (2016), however, the procedure of Montanaro is not hybrid, and uses a quantum counting algorithm to determine whether a valid solution exists at a deeper level of the tree. As has been shown by Creemers and Pérez (2023), however, quantum counting algorithms are dominated by GUM when verifying whether a valid solution exists.

To further illustrate the dynamics of HBB, we use HBB to solve the example binary knapsack problem. The steps of the procedure are illustrated in Figure 9. In this example, we assume $\omega = 1$. To initialize the procedure, we let $\mathbf{x}^* = \emptyset$, $V^* = 0$, and enter a first recursion with current solution $\mathbf{x} = \emptyset$, value $V = 0$, remaining weight $R = W = 4$, and $i = 1$. In this first recursion, a classical procedure should not be used (i.e., $n - i \geq \omega$), and sufficient weight remains for item $i = 1$ to be added (i.e., $w_1 \leq R$). As a result, we update the current partial solution, and let $x_1 = 1$, $\mathbf{x}_{11} = \{1\}$, $V' = 3$, and $R' = 2$. We also update the best-found solution and let $\mathbf{x}^* = \{1, 0, 0\}$ and $V^* = 3$. Next, we use GUM to measure a solution $\mathbf{x}_{(i+1)n} \in \Omega_{(i+1)n}$ that has a value $V_{(i+1)n}$ larger than $V^* - V'$ (i.e., a value larger than 0) and a weight $W_{(i+1)n}$ that does not exceed R'. After adding item 1, two items are left, and $\mathbf{x}_{(i+1)n} = \{0, 1\}$ is the only valid solution (i.e., $\mathbf{x}^*$ can only be improved if we add item 3). In a system that has two items (i.e., qubits) and only one valid solution, GUM has a 100 percent probability to measure the valid solution after performing $I(2, 4) = 1$ Grover iteration. As a result, after calling GUM, we update the best-found solution, and let $\mathbf{x}^* = \{1, 0, 1\}$ and $V^* = 5$. After updating the best-found solution, we enter a second recursion. In this recursion, $(n - i)$ is still smaller than ω, and as such, we do not use a classical procedure to try and improve the current solution. We are, however, also unable to add item 2 to the existing solution (that already contains item 1; $w_2 > R$). As a result, we enter a new recursion without adding item 2. In this recursion, $(n - i) < \omega$, and we use a classical procedure

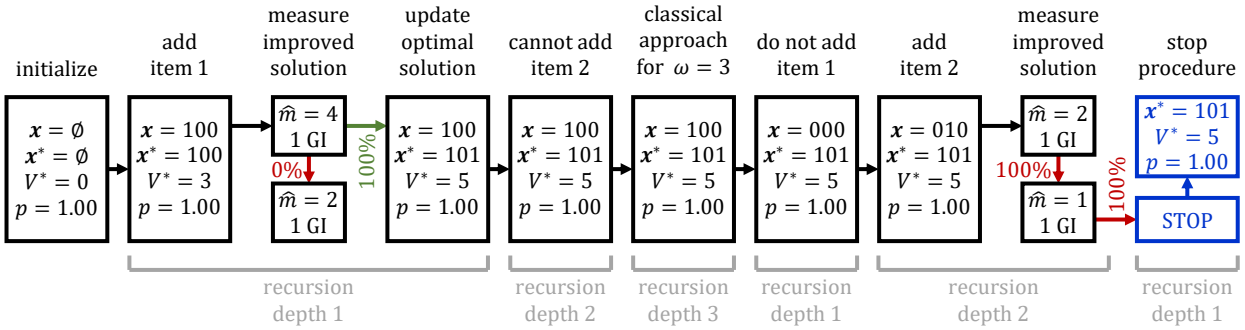
Algorithm 3: Hybrid branch-and-bound procedure to solve the binary knapsack problem using GUM.

initialize $\mathbf{x}^* = \emptyset$, $V^* = 0$, and start procedure HBB$(\emptyset, 0, W, 1)$;
Procedure HBB$(\mathbf{x}_{1(i-1)}, V, R, i)$
  if $n - i < \omega$ **then**
    use classical algorithm to find $\mathbf{x}^\star_{in} \in \Omega_{in}$ that has highest value $V^\star_{in} > V^* - V$ and $W^\star_{in} \leq R$;
    if $\mathbf{x}^\star_{in} \neq \emptyset$ **then**
      update best-found solution $\mathbf{x}^* = \mathbf{x}_{1(i-1)} \cup \mathbf{x}^\star_{in}$ and its value $V^* = V + V^\star_{in}$;
  else
    if $w_i \leq R$ **then**
      Let $x_i = 1$, $\mathbf{x}_{1i} = \mathbf{x}_{1(i-1)} \cup \{x_i\}$, $V' = V + v_i$, and $R' = R - w_i$;
      if $V' > V^*$ **then**
        Update best-found solution $\mathbf{x}^* = \mathbf{x}_{1i} \cup \{\mathbf{0}\}$ and its value $V^* = V'$;
      Use GUM$(n - i)$ to measure $\mathbf{x}_{(i+1)n} \in \Omega_{(i+1)n}$ that has value $V_{(i+1)n} > V^* - V'$ and $W_{(i+1)n} \leq R'$;
      if $\mathbf{x}_{(i+1)n} \neq \emptyset$ **then**
        update best-found solution $\mathbf{x}^* = \mathbf{x}_{1i} \cup \mathbf{x}_{(i+1)n}$ and its value $V^* = V' + V_{(i+1)n}$;
        HBB$(\mathbf{x}_{1i}, V', R', i + 1)$;
    Let $x_i = 0$, $\mathbf{x}_{1i} = \mathbf{x}_{1(i-1)} \cup \{x_i\}$, and run HBB$(\mathbf{x}_{1i}, V, R, i + 1)$;

to determine whether a partial solution can still be found that allows to improve the best-found solution. As the optimal solution has already been found, no improvement is possible, and we exit this recursion. After ascending to recursion depth 2, we again ascend one level, and end up at recursion level 1, in which we enter a new recursion where item 1 is not added. In this recursion, the current value equals $V = 0$ and the remaining weight capacity is $R = 4$. As a result, item 2 can be added, we update the current solution, and use GUM to measure a solution $\mathbf{x}_{(i+1)n} \in \Omega_{(i+1)n}$ that has a value bigger than $V^* - V' = 4$. No such solution can be found, and therefore, after performing two Grover iterations (one for $\widehat{m} = 2$ and one for $\widehat{m} = 1$), we complete the procedure with a 100 percent probability to find the optimal solution. In total, to find the optimal solution, HBB requires 1 Grover iteration in which a solution of two items is evaluated and 2 Grover iterations in which a solution of a single item is evaluated (in contrast to BSP which requires at most 13 Grover iterations in which a solution of three items is evaluated; note that it requires more operations to evaluate a solution that has more items).

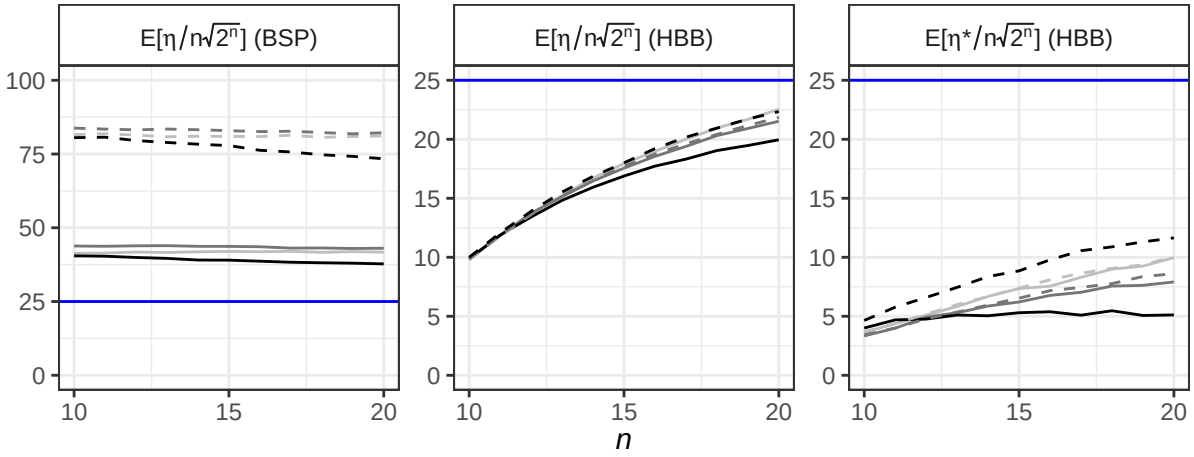
To assess the performance of HBB, we use Monte Carlo simulation to simulate the execution of HBB 1000 times for each of the problem instances defined in the previous section for $n : n \in \{10, 11, \ldots, 19, 20\}$ (resulting in 66000 instances of the binary knapsack problem). In the experiment, we let $\omega = 5$ (i.e., if 5 items remain, a classical approach is used to determine which partial solution best improves the current solution; a value $\omega = 5$ is chosen because Grover's algorithm is sufficiently accurate when at least 6 items are considered). Note that, in the discussion of the results that follows, we omit the (insignificant) number of operations required by the classical approach. The results of the experiment are presented in Figure 10. Figure 10 shows the expected number operations divided by $n\sqrt{2^n}$ (denoted $E[\eta/n\sqrt{2^n}]$; note that we divide

Figure 9. Illustration of the dynamics of procedure HBB when solving the example binary knapsack problem.



by $n\sqrt{2^n}$ to show that HBB requires $O(\eta/n\sqrt{2^n})$ operations, where η is a constant) when solving the binary knapsack problems using HBB (center panel). For comparison, we also include the performance of BSP (left panel). It is obvious that HBB significantly outperforms BSP. In addition, it seems that HBB can solve instances that have a larger range of values (i.e., instances of sets 4, 5, and 6) without extra effort (this in contrast to BSP that uses a binary search procedure to identify the optimal value; a larger range of values results in more operations). For both procedures, however, we observe (slight) differences in the required number of operations to solve instances of different sets. To further explore these differences, we refer to Table 6 that reports the expected number of times that a feasible solution value occurs over all instances in a set. For instance, for set 3, Table 6 shows that a feasible solution value is expected to occur 252.58 times for an instance that has $n = 20$ items (i.e., given a feasible solution value V, we expect to find 252.58 feasible solutions that have the same value V). In contrast, for set 4, a feasible solution value only is expected to occur 1.06 times for an instance that has $n = 20$ items. If more solutions have a particular value V, more solutions can be removed from the search space if GUM can show that a solution exists that has a value higher than V (i.e., it is easier to find the best solution in a set of solutions that have 2 values that occur 500 times than in a set of solutions that have 1000 values that only occur once). As a result, for BSP we see small differences between sets 1 and 2, and between sets 4, 5, and 6 (with slightly increasing differences for instances of set 6 if n is large). For HBB, that is not impacted by an increase in the range of values, we see almost no difference between sets 1 and 4 (instances that have uncorrelated weights/values), and between sets 2 and 5 (instances that have correlated weights/values). There is, however, a slight difference between instances that have uncorrelated weights/values (sets 1 and 4) and instances that have correlated weights/values (sets 2 and 5). This illustrates that, rather than the range of values, the structure of the instance can have an impact on the performance of HBB. We also observe that uncorrelated/correlated instances (sets 1, 2, 4, and 5) differ from profit-ceiling instances that are both the most easy to solve (set 3) and the most difficult to solve (set 6). Here again, the instances of set 3 are easiest to solve because they have a significantly higher frequency of feasible solution values. Set 6, on the other hand, may be the hardest to solve because the branch-and-bound procedure is less capable of exploiting the structure of profit-ceiling instances (when compared to uncorrelated/correlated instances, profit-ceiling instances have items that have a ratio of value over weight that is always close to 1). Notwithstanding the differences among sets, it appears that, for all sets, the expected number of operations (divided by $n\sqrt{2^n}$) converges to some constant (note that this also has been verified for some sets for values

Figure 10. Expected number of operations divided by $n\sqrt{2^n}$ required to solve a binary knapsack problem using procedures BSP and HBB, as well as the expected number of operations divided by $n\sqrt{2^n}$ required to find an optimal solution for the first time using HBB, when solving 1000 binary knapsack problems for each problem set and for each value of n : $10 \leq n \leq 20$ (lines of sets 1 and 4 are light gray, sets 2 and 5 are dark gray, sets 3 and 6 are black, sets 1, 2, and 3 are full, and sets 4, 5, 6 are dashed). The blue line indicates a value of 25 on the vertical axis.



of n larger than 20). Therefore, we conclude that HBB has expected runtime $O(n\sqrt{2^n})$, which is equivalent to the performance of the best classical algorithms that solve the binary knapsack problem. Last but not least, Figure 10 also reports the expected number of operations required to find an optimal solution for the first time (denoted $E[\eta^*/n\sqrt{2^n}]$; right panel). We observe that HBB can find an optimal solution for the first time in $O(n\sqrt{2^n})$ operations. The difference between the right and center panel, however, also shows that, once an optimal solution has been found, a significant amount of time is spent to prove that this solution is optimal (i.e., in different nodes of the search tree, GUM needs to evaluate all values of $\hat{m}$ in order to be sure that no valid solution can be found that improves the optimal solution). This seems to be the main bottleneck of HBB. In the upcoming section, we try to resolve this issue, and present a procedure that minimizes the time that is spent to show that a solution is optimal.

8. Random-ascent procedure

In the previous section, we presented a hybrid branch-and-bound procedure (HBB) that is able to solve a binary knapsack problem that has n items using only $O(n\sqrt{2^n})$ operations. Most of these operations, however, are spent on proving that the optimal solution cannot be improved. To minimize the time spent on proving that a solution is optimal, we propose another quantum algorithm. This simple, yet elegant, procedure requires only one call to GUM once an optimal solution has been found. The procedure can be seen as a random-ascent procedure that achieves a random improvement of the best-found solution value in each iteration. In what follows, we use this procedure to solve the binary knapsack problem. Algorithm 4 (hereafter referred to as RAP) provides an outline of the procedure. After initializing $x^* = \emptyset$ and $V^* = 0$, RAP enters a first iteration in which GUM is used to measure a solution that has a value $V' > V^*$ and weight $W' \leq W$. If such a solution is found, we update the best-found solution, and start a new iteration. This process continues until an optimal

Table 6. Average expected number of times that a feasible solution value occurs for each problem set and for each value of n : $10 \leq n \leq 20$.

n	set 1	set 2	set 3	set 4	set 5	set 6
10	1.10	1.10	1.41	1.00	1.00	1.00
11	1.19	1.20	1.81	1.00	1.00	1.00
12	1.38	1.39	2.61	1.00	1.00	1.00
13	1.75	1.78	4.13	1.00	1.00	1.00
14	2.49	2.57	6.96	1.00	1.00	1.00
15	3.92	4.10	12.17	1.00	1.00	1.01
16	6.60	6.95	21.78	1.00	1.01	1.02
17	11.54	12.20	39.62	1.01	1.01	1.03
18	20.69	21.92	72.88	1.02	1.02	1.06
19	37.72	40.03	135.20	1.03	1.03	1.12
20	69.57	73.95	252.58	1.06	1.07	1.25

solution has been found. If we are lucky, we find an optimal solution after one iteration. If we're unlucky, it may take $S \leq 2^n$ iterations (where S is the number of feasible solution values). In general, however, we expect to find an optimal solution after $(\lceil \log_2(S) \rceil + 1)$ iterations. In addition, in the first iterations, we may expect `GUM` to find a solution using few Grover iterations (as there are still many solutions that can improve the best-found solution at that point). In later iterations (if few valid solutions remain), `GUM` has to evaluate more (increasingly smaller) values of $\widehat{m}$ that require (increasingly larger) number of Grover iterations. If, eventually, an optimal solution has been found, `GUM` is called one last time to verify that no solution can be found that still improves the optimal solution.

Algorithm 4: Random-ascent procedure to solve the binary knapsack problem using `GUM`.

initialize $\mathbf{x}^* = \emptyset$, $V^* = 0$, and start procedure `RAP(`n`)`;
Procedure `RAP(`n`)`
  do
    Use `GUM(`n`)` to measure $\mathbf{x}'$ that has value $V' > V^*$ and $W' \leq W$;
    if $\mathbf{x}' \neq \emptyset$ **then**
      update best-found solution $\mathbf{x}^* = \mathbf{x}'$ and its value $V^* = V'$;
  while $\mathbf{x}' \neq \emptyset$;

To further illustrate the dynamics of `RAP`, we use it to solve the example binary knapsack problem. The steps of the procedure are illustrated in Figure 11. After initializing, we use `GUM` to measure any of the four feasible solutions that have a value $V > 0$. There is a 6.25 percent probability that no valid solution was measured (leading us to conclude that there is no feasible solution). There is, however, also a 93.75 percent probability that a valid solution was measured. In this case, there are 4 options. First, we may measure the optimal solution itself (with a probability of 25 percent). After measuring the optimal solution, we perform one more run of `GUM` to show that no solution can still be found that improves the optimal solution. Second, we may measure solution $\mathbf{x} = \{1, 0, 0\}$ that has $V = 3$. In this case, we perform another iteration in which we run `GUM` again and have a 99.99 percent probability to measure the only solution that can still improve a value of $V = 3$ (i.e., the optimal solution that has $V = 5$). After measuring the optimal solution, we

perform one more run of GUM to verify that the optimal solution has been found. Third, we may measure solution $\mathbf{x} = \{0, 0, 1\}$ that has $V = 2$. In this case, we have a 100 percent probability to measure one of two solutions that have a value $V > 2$. If the optimal solution is measured, we run GUM to show that no more improvement can be made. If, on the other hand, we measure solution $\mathbf{x} = \{1, 0, 0\}$, we need to run GUM again to measure the optimal solution and once more to prove that there is no solution that can still improve the optimal solution. Fourth, we may measure solution $\mathbf{x} = \{0, 0, 1\}$ that has $V = 1$. In this case, all the aforementioned scenarios are possible after running GUM again. In the end, RAP has a 93.20 percent probability to measure the optimal solution (see also the blue boxes in Figure 11).

In the worst case, there are $S = 2^n$ feasible solution values, and we expect to perform $\lceil \log_2(S) \rceil + 1 = (n + 1)$ iterations of GUM. In each iteration $i : 0 \leq i \leq n$, we expect that $m = {}^{2^n}/_{2^i} = 2^{(n-i)}$ feasible solution values remain, and that GUM performs $\sum_{j=0}^{n} (1-\phi_{i-1}) P(n, 2^{(n-i)}, 2^{(n-j)}) I(n, 2^{(n-j)})$ Grover iterations, where ϕ_i is the probability of having measured a valid solution after i GUM iterations and $P(n, m, I)$ is the probability to measure one of the m valid solutions in a set of 2^n solutions after performing I Grover iterations (see Appendix A for a simple procedure to calculate $P(n, m, I)$). ϕ_i may be obtained through the following recursive relationship:

$$\phi_i = \phi_{i-1} + (1 - \phi_{i-1}) P(n, m, I(n, 2^{(n-i)})), \tag{7}$$

where $\phi_{-1} = 0$. We expect to find an optimal solution after $(n + 1)$ iterations of GUM, after which we call GUM one last time to prove that the optimal solution cannot be improved. In this last run of GUM, we require $\sum_{j=0}^{n} I(n, 2^{(n-j)})$ Grover iterations. As explained earlier, each Grover iteration requires two calls to a function $f_{\mathbf{x}}$. In addition, we expect to validate at most $(n + 2)$ measured solutions, requiring another $(n + 2)$ calls to function $f_{\mathbf{x}}$. Therefore, in the worst case, we expect that RAP requires $\sum_{j=0}^{n} 2I(n, 2^{(n-j)}) + (n + 2) + \sum_{i=0}^{n} \sum_{j=0}^{n} \left(1 - \phi_{i-1}\right) P(n, 2^{(n-i)}, I(n, 2^{(n-j)})) 2I(n, 2^{(n-j)})$ calls to function $f_{\mathbf{x}}$. Figure 12 illustrates the worst-case number of expected calls required by RAP divided by $\sqrt{2^n}$ for $n : 1 \leq n \leq 60$. We observe that the worst-case number of expected calls converges to $13.72\sqrt{2^n}$, implying that: (1) RAP has worst-case expected runtime $O(\mu\sqrt{2^n})$ for solving any discrete optimization problem that has n binary decision variables, and (2) RAP can solve the binary knapsack problem using $O(n\sqrt{2^n})$ operations (in the worst case; thereby matching the performance of the best classical algorithms).

To verify the effective runtime of RAP, we use Monte Carlo simulation to simulate its execution 1000 times for each of the aforementioned 66000 problem instances. The results are summarized in Figure 13 that shows the expected number of operations (divided by $n\sqrt{2^n}$) when solving the binary knapsack problems using HBB and RAP (top left and right panels, respectively). We observe that RAP easily outperforms HBB, and that its performance remains well below its worst-case performance (indicated by the red line in the right top panel). Figure 13 also shows the expected number of operations (divided by $n\sqrt{2^n}$) that are required to find an optimal solution for the first time. Here also, we see that RAP (right bottom panel) outperforms HBB (left bottom panel). Before we conclude, however, that RAP dominates HBB, we first investigate whether the performance of both procedures can be improved.

Figure 11. Illustration of the dynamics of RAP when solving the example binary knapsack problem.

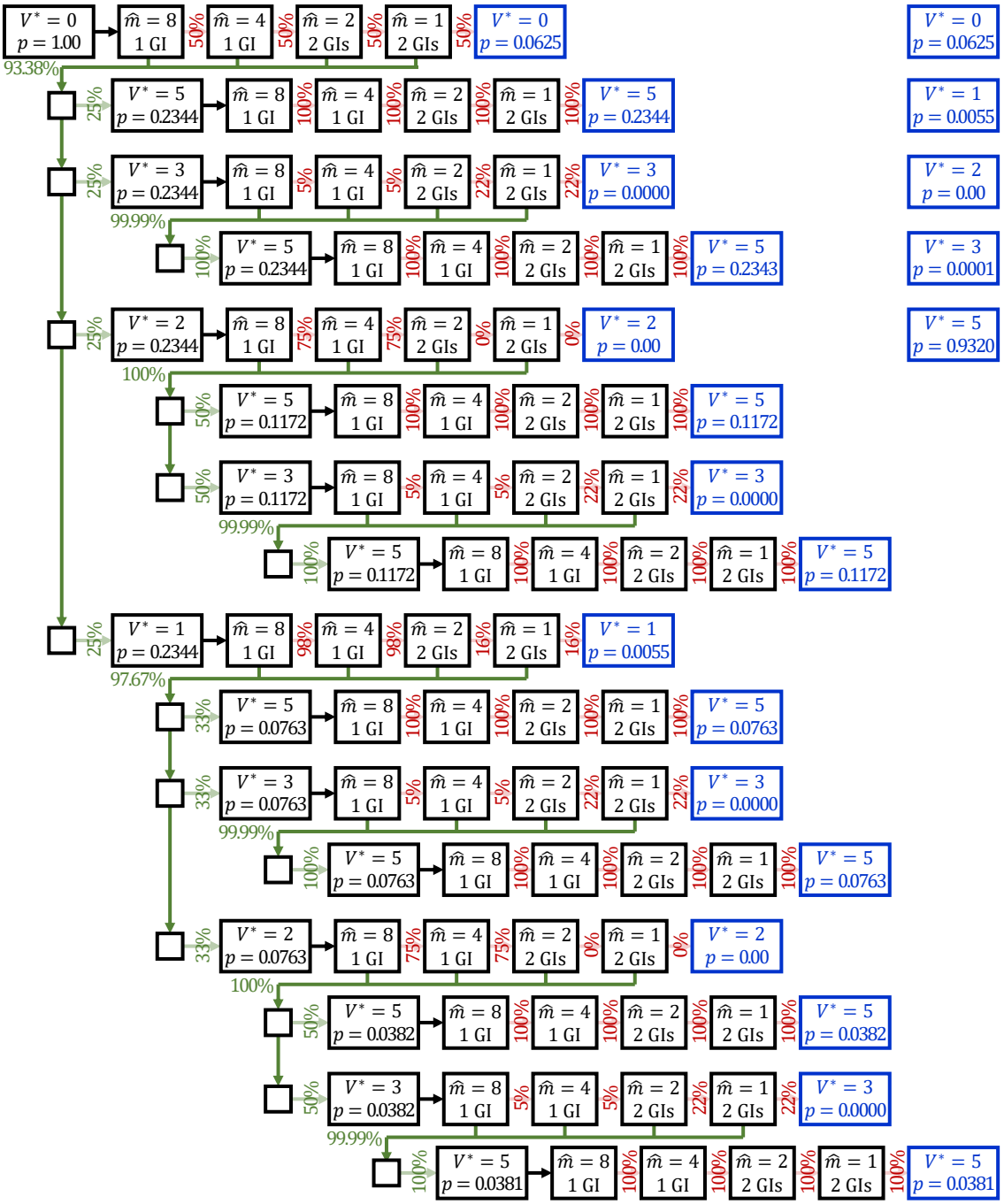


Figure 12. Worst-case expected number of operations divided by $n\sqrt{2^n}$ required by procedure `RAP` for each value of n : $1 \leq n \leq 60$.

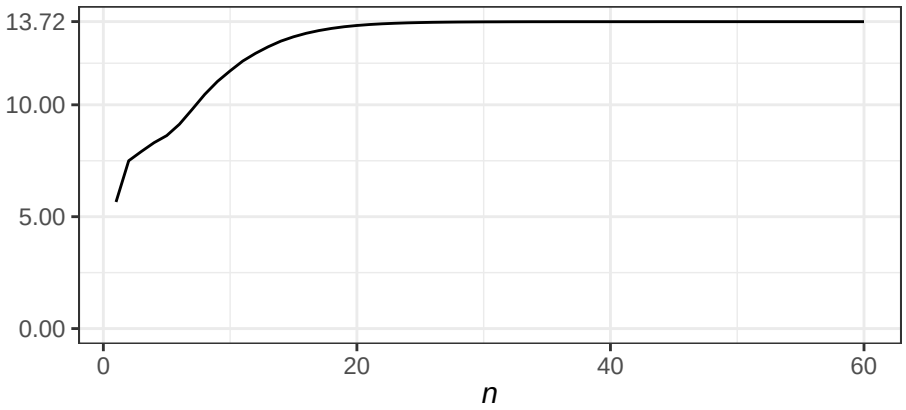
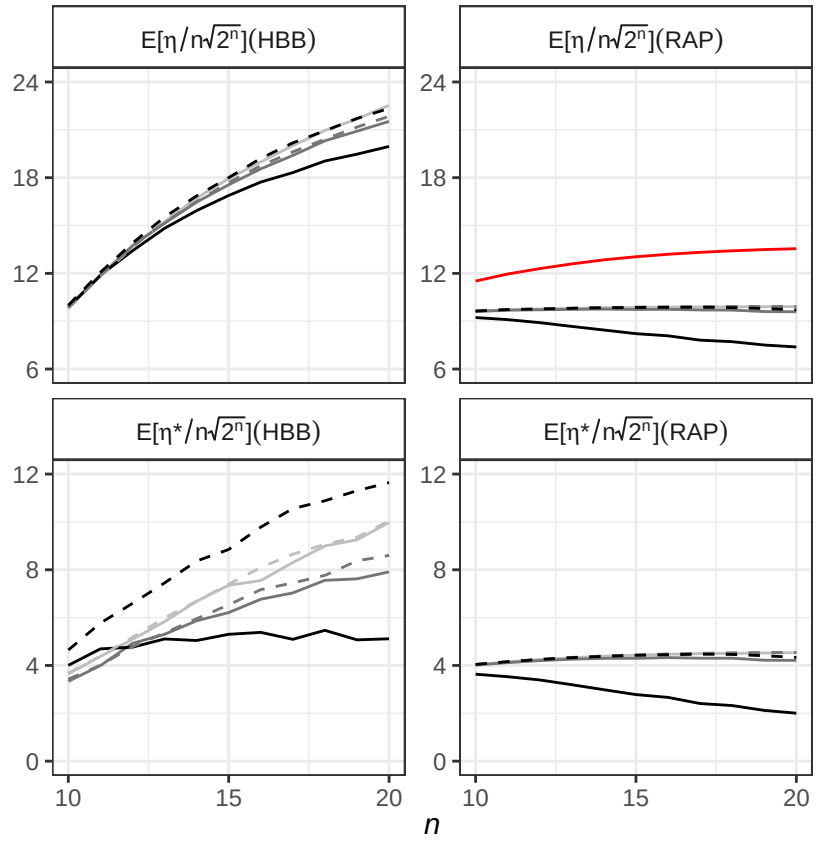


Figure 13. Expected number of operations divided by $n\sqrt{2^n}$ required to solve a binary knapsack problem using `HBB` and `RAP`, as well as expected number of operations divided by $n\sqrt{2^n}$ required to find an optimal solution for the first time using both procedures, when solving 1000 binary knapsack problems for each problem set and for each value of n : $10 \leq n \leq 20$ (lines of sets 1 and 4 are light gray, sets 2 and 5 are dark gray, sets 3 and 6 are black, sets 1, 2, and 3 are full, and sets 4, 5, 6 are dashed). The worst-case expected performance of procedure `RAP` is indicated by the red line (see also Figure 12).



9. Improving procedures HBB and RAP

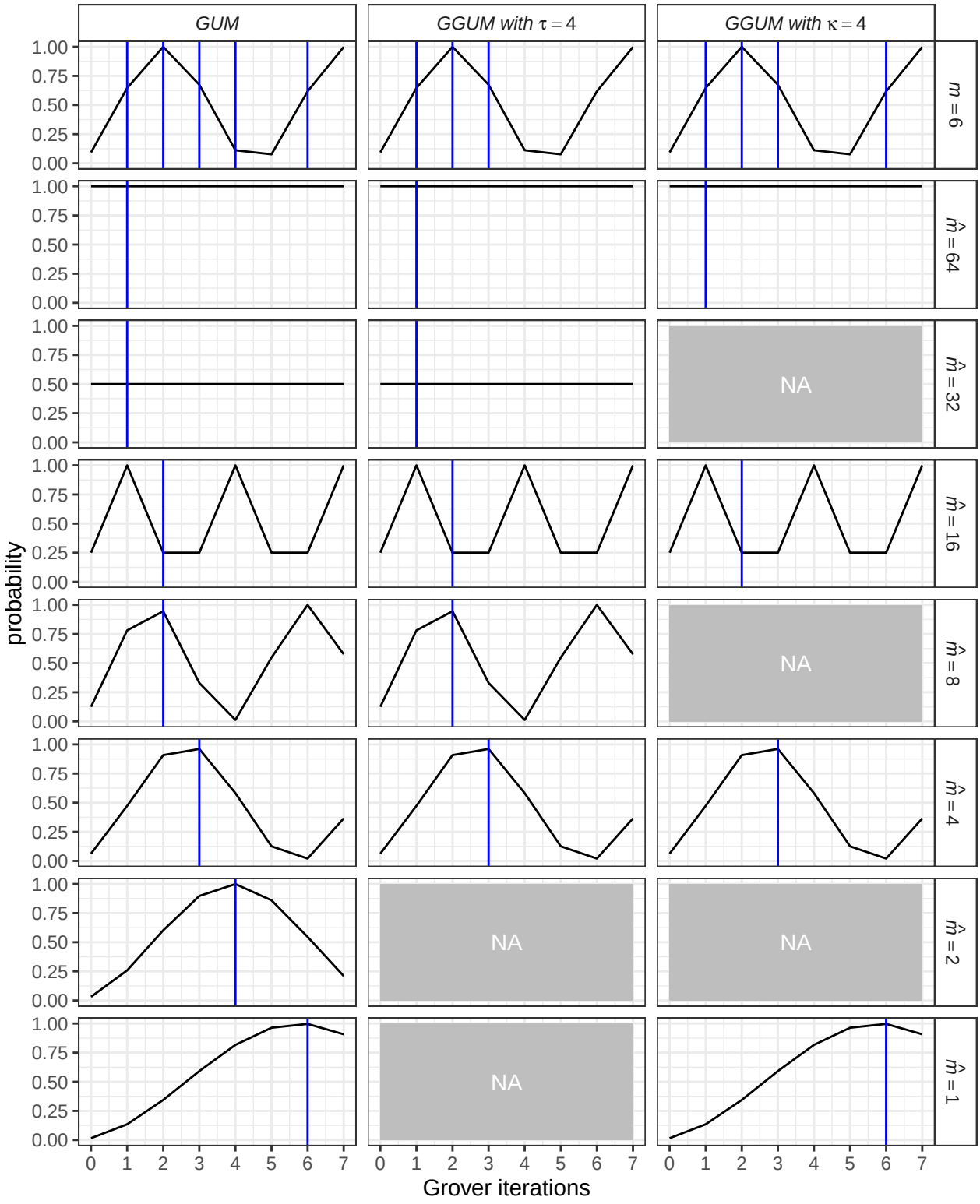
Both HBB and RAP can be improved in several ways. First, a heuristic solution procedure can be used to provide a better initial solution (resulting in less recursions/iterations to be performed by both procedures, respectively). Second, GUM lies at the core of the performance of HBB and RAP. Therefore, if we can improve GUM, we also improve HBB and RAP. Note that, as the true number of valid solutions (i.e., m) is unknown, GUM takes a sample of assumed numbers of valid solutions (i.e., $\widehat{m}$) by initializing $\widehat{m}$ as the total number of solutions, and by letting $\widehat{m} = \widehat{m}/2$ in subsequent iterations. If no solution is found after evaluating $\widehat{m} = 1$, the procedure stops. Better sampling methods, however, may exist (see, for example, Boyer et al. (1996) for an alternative sampling method). In this section, we investigate alternative sampling methods, and generalize GUM by: (1) dividing $\widehat{m}$ by κ rather than by 2, and (2) stopping the procedure once $\widehat{m} \leq \tau$. Figure 14 illustrates the sample of values of $\widehat{m}$ (and the corresponding number of Grover iterations) that is used in a system that has $n = 6$ qubits and $m = 6$ valid solutions: (1) in case of regular GUM (left column), (2) if we let $\tau = 4$ (middle column), and (3) if we let $\kappa = 4$ (right column). For instance, if $\tau = 4$, we do not evaluate values $\widehat{m} = 2$ and $\widehat{m} = 1$. If, on the other hand, $\kappa = 4$, we skip $\widehat{m} = 32$, $\widehat{m} = 8$, and $\widehat{m} = 2$. If less values of $\widehat{m}$ are evaluated, less Grover iterations are required, however, it also becomes less likely that we measure a valid solution (in the three cases, after evaluating all values of $\widehat{m}$, we have a probability of 1.00, 1.00, and 0.99 to measure one of the $m = 6$ valid solutions, respectively). In addition, as another improvement, note that, currently, HBB and RAP do not use any information of past runs of GUM. If a solution was found assuming a particular value of $\widehat{m}$, however, it makes sense to assume that in a subsequent (more restricted) search there are at most $\widehat{m}$ valid solutions (i.e., it does not make sense to once more reinitialize $\widehat{m}$ as the total number of solutions; we can use the value of $\widehat{m}$ for which a solution was found in a previous search to start searching in a subsequent, more restricted search). These ideas are reflected in Algorithm 5 (referred to as GGUM in what follows) that generalizes GUM.

Algorithm 5: Generalized procedure that uses Grover's algorithm to find a valid solution when the number of valid solutions is unknown in a system that has n binary decision variables, and where $f_{\mathbf{x}}$ returns 1 if solution $\mathbf{x}$ is valid (and 0 otherwise).

```
procedure GGUM(n, m̂)
    do
        perform I(n, m̂) iterations of Grover's algorithm equipped with f_x;
        measure basis state |x⟩;
        if f_x = 1 then
            return x and m̂;
        else if m̂ > τ then
            m̂ = m̂/κ;
        else
            return ∅;
    while true;
```

Next to these (potential) improvements, HBB may also benefit from classical techniques to improve the

Figure 14. Sample of values of $\hat{m}$ (and corresponding number of Grover iterations) that are used in case of regular `GUM`, `GGUM` with $\tau = 4$, and `GGUM` with $\kappa = 4$. Each graph presents the probability to measure a valid solution in a system that has $n = 6$ qubits if $m = 6$ (top row), $\hat{m} = 64$ (second row), $\hat{m} = 32$ (third row), $\hat{m} = 16$ (fourth row), $\hat{m} = 8$ (fifth row), $\hat{m} = 4$ (sixth row), $\hat{m} = 2$ (seventh row), and $\hat{m} = 1$ (eight row). In rows 2 - 8, the blue line represents the number of Grover iterations given by $I(n, \hat{m})$. In the top row, the blue lines correspond to the sample of values of $\hat{m}$ used by each approach to measure any of the $m = 6$ valid solutions.



performance of a branch-and-bound procedure. For instance, when solving the binary knapsack problem, items can be sorted by their ratio of value over weight. In addition, Algorithm 6 can be used to calculate an upper bound on the maximum value that can still be obtained in any node at level i (hereafter referred to as `UBi`; see also Dantzig (1957)). If this upper bound does not exceed the best-found solution, we can fathom the node. Note that, in contrast to `HBB`, `RAP` cannot benefit from this upper-bound procedure.

Algorithm 6: Procedure that calculates an upper bound on the value of the current knapsack solution (that has value V and remaining weight capacity R) in a node at level i of the search tree. Note that we assume that items are sorted by their ratio of value over weight.

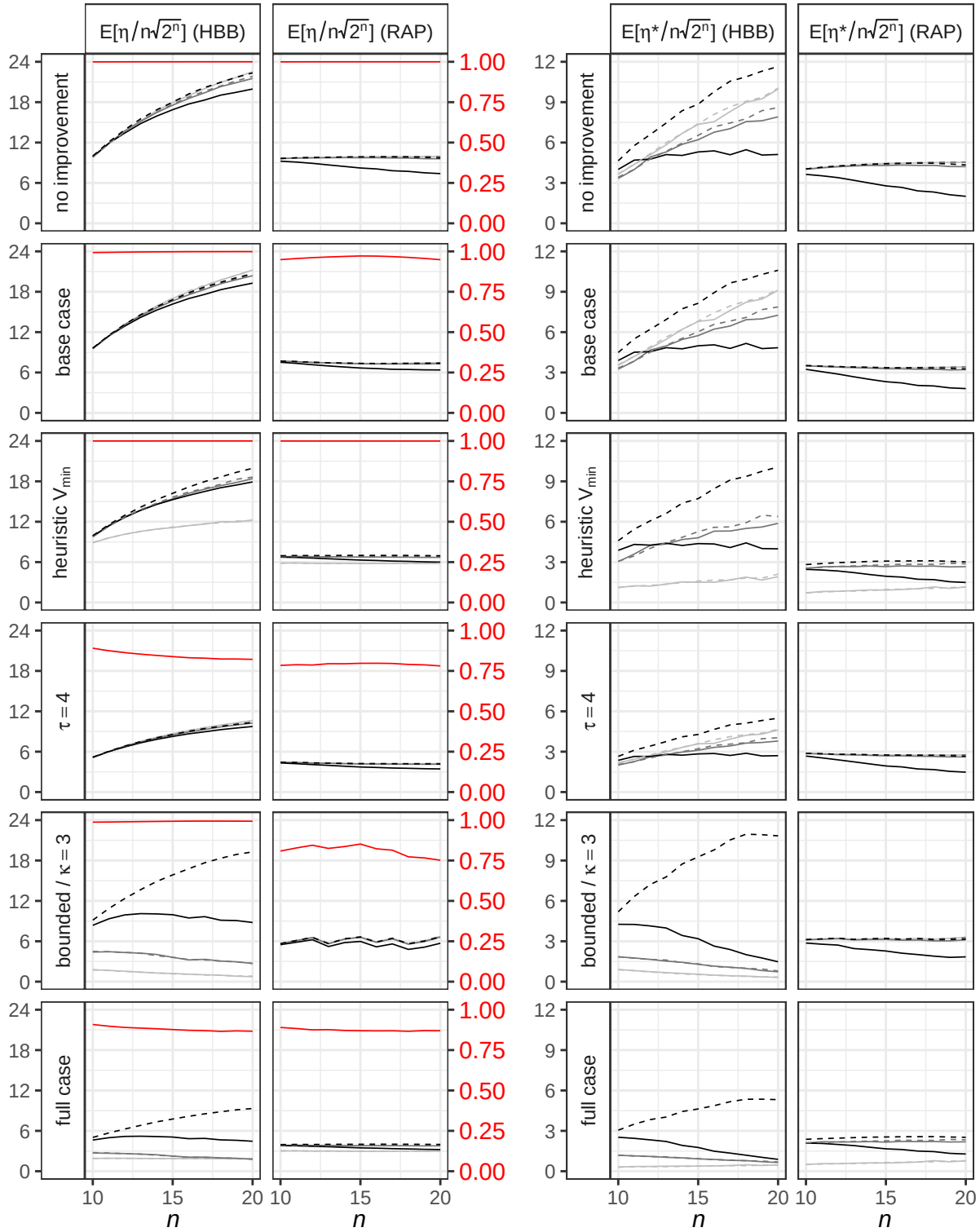
```
procedure UBi(V, R, i)
    for j = i to n do
        if w_j ≤ R then
            let V = V + v_j and R = R − w_j;
        else
            let V = V + (R / w_j) v_j;
            break;
    return V;
```

In what follows, we perform a computational experiment in which we solve 66000 instances of the binary knapsack problem, and assess the impact of:

- using information of the past run of `GGUM` to determine the initial value of $\widehat{m}$ in the subsequent run of `GGUM` (referred to as the "base case"; all improvements mentioned below depart from this base case).
- using a heuristic procedure to calculate a lower bound that serves as an initial solution. The heuristic procedure first sorts items based on their ratio of value over weight (higher ratios are ranked first). Next, we initialize an empty knapsack, and add items in order of the list. If an item cannot be added (because its weight exceeds the remaining weight capacity), we move to the next item in the list. The procedure stops if all items in the list have been processed.
- assuming a value of $\tau = 4$ (i.e., $\widehat{m} = 1$ and $\widehat{m} = 2$ are not evaluated by `GGUM`).
- assuming a value of $\kappa = 3$ (i.e., we evaluate $\widehat{m} = 2^n$, $\widehat{m} = {}^{2^n}/_{3}$, $\widehat{m} = {}^{2^n}/_{9}$, etc.).
- using upper bound `UBi` to fathom nodes when using `HBB`.

The results of the experiment are summarized in Figure 15. For different scenarios, Figure 15 reports the expected number of operations (divided by $n\sqrt{2^n}$) required by `HBB` and `RAP` (left two columns) as well as the expected number of operations (divided by $n\sqrt{2^n}$) required to find an optimal solution for the first time (right two columns). Figure 15 also shows the average probability to find an optimal solution for each of the scenarios/procedures (red line in the left two columns; note that the expected number of operations to find an optimal solution for the first time only is calculated for those instances for which an optimal solution was

Figure 15. Expected operations (divided by $n\sqrt{2^n}$) to solve a binary knapsack problem using `HBB` and `RAP` as well as the expected operations (divided by $n\sqrt{2^n}$) to find an optimal solution for the first time, when solving 1000 knapsack problems for each problem set, for various values of n, and for various scenarios (lines of sets 1 and 4 are light gray, sets 2 and 5 are dark gray, sets 3 and 6 are black, sets 1, 2, and 3 are full, and sets 4, 5, 6 are dashed). A red line shows the average probability to identify an optimal solution.



found). For comparison, we also included the results of HBB and RAP without improvements (first row; these results are identical to those presented in Figure 13).

A first conclusion that can be drawn by observing Figure 15 is that initializing $\widehat{m}$ as the last value of $\widehat{m}$ in the previous run (base case; reported in row 2) has a significant impact on the number of required operations without significantly reducing the probability to measure an optimal solution. Next, even though the use of a lower bound (row 3) has a positive impact, the effect is limited except for sets 1 and 4 (where items have uncorrelated values and weights). This can be explained by the fact that items in sets 2 and 5 have similar ratios (values and weights are correlated) and items in sets 3 and 6 have ratios close to 1 (item i has value $v_i = 3\lceil w_i/3 \rceil$)). As a result, for these sets, our lower bound (based on the ratio of value over weight) may not be very effective. In row 4, on the other hand, we see that letting $\tau = 4$ once more has a significant effect on the number of required iterations without impacting the probability to measure an optimal solution too much. In contrast, in row 5, we evaluate the impact of letting $\kappa = 3$ for RAP (second and fourth column). Even though we are able to reduce the number of required operations, we also observe that the probability to measure an optimal solution reduces as n increases. In other words, with $\kappa = 3$, the sample of values of $\widehat{m}$ becomes too sparse to still guarantee that GGUM can still measure a valid (and hence optimal) solution. In row 5, we also illustrate the impact of upper bound UBi on the performance of HBB. We observe a significant reduction of the number of operations required to find an optimal solution (first column) and the number of operations required to find an optimal solution for the first time (third column). The impact is most outspoken for instances where the structure of the problem can be exploited (i.e., instances with uncorrelated and correlated values and weights). For sets 3 and 6, the impact is less clear. For instances in set 6, the effectiveness of UBi is limited by the fact that items have a ratio of value over weight that is close to 1. For set 3, this effect is offset by the fact that instances in set 3 have feasible solutions values that have a high frequency (see also Table 6). Last but not least, in row 6 we evaluate the impact of all improvements (except $\kappa = 3$ and upper bound UBi in the case of RAP; referred to as "full case" in Figure 15) on the performance of HBB and RAP. Here we observe that both procedures have roughly the same probability to find an optimal solution, however, RAP no longer dominates HBB. Only when solving instances of set 6, RAP still has the upper hand (i.e., RAP may only be able to outperform HBB if HBB cannot exploit the structure of the problem). Perhaps more importantly, row 6 also illustrates that HBB and RAP may require (far) less than $O(n\sqrt{2^n})$ operations to find an optimal solution for the first time. This indicates that they may be used as efficient heuristics to obtain (near-) optimal solutions by observing the best-found solution after a given time.

10. Conclusions

In this paper, we investigate the potential of quantum computing to solve discrete optimization problems. for this purpose, we first introduce the fundamentals of quantum computing as well as the Deutsch algorithm. Even though the Deutsch algorithm solves a problem that has no practical use, it is of particular interest because it serves as a stepping stone towards Grover's algorithm. Arguably, Grover's algorithm is the most important quantum algorithm. It can be used to perform an unstructured search and achieves a quadratic

speedup when compared to classical algorithms. As a search algorithm, it can be used to solve discrete optimization problems by identifying valid solutions (i.e., feasible solutions that have an objective value of at least V; assuming we consider a maximization problem). To identify one of the m valid solutions in a set of 2^n solutions, Grover's algorithm needs $O(\sqrt{m^{-1}2^n})$ calls to a function $f_{\mathbf{x}}$ that returns 1 if solution $\mathbf{x}$ is valid (and 0 otherwise).

When using Grover's algorithm to solve discrete optimization problems, we essentially face two problems: (1) we don't know the number of valid solutions that have a value of at least V, and (2) we don't know the optimal solution value V^*. To resolve the first problem, we use a procedure (GUM) that samples different values of $\widehat{m}$ (i.e., the assumed number of valid solutions) and that is almost certain to return a valid solution (if it exists). We show that GUM requires at most $O(\sqrt{2^n})$ calls to function $f_{\mathbf{x}}$ if we are solving a discrete optimization problem that has n binary decision variables. To resolve the second problem, we first adopt a binary search procedure (BSP) that evaluates L values objective values V (where L is a logarithmic function of the bounds on the optimal solution value V^*). Next, with both problems resolved, we use BSP and illustrate how to effectively solve the binary knapsack problem. In addition, we show how to assess the performance of BSP and perform a computational experiment in which 108000 binary knapsack problems are solved. Our results indicate that BSP requires $O(nL\sqrt{2^n})$ operations. The best classical algorithms, however, require $O(\sqrt{2^n})$ operations. As such, BSP is unable to match the performance of the best classical algorithms when solving the binary knapsack problem.

To improve this result, we propose two new procedures that also rely on GUM to find a valid solution: a hybrid branch-and-bound procedure (HBB) and a random-ascent procedure (RAP). To solve any discrete optimization problem that has n decision variables, HBB uses a depth-first search strategy to explore a tree that has n levels. At level $i : 1 \leq i \leq n$, we visit nodes that represent partial solutions where values have been assigned to decision variables $j < i$. If we are currently in a node at level i, and x_i is the (valid) value that has been assigned to decision variable i, HBB uses GUM to determine whether we can find a valid assignment for decision variables $j > i$ such that the best solution that was found so far can still be improved. Only if such a solution can be found, HBB explores the branch where decision variable i is assigned value x_i. Because Grover's algorithm (and hence GUM) is less accurate when finding solutions for smaller problems, we propose a hybrid approach, and use a classical algorithm to solve partial problems that consider ω decision variables or less. In addition, note that HBB can exploit classical techniques to improve the performance of a branch-and-bound procedure (e.g., dominance rules and/or bounding procedures). In contrast, RAP, cannot exploit such techniques, and is a far more simple procedure that iteratively uses GUM to draw a "random" solution from the set of solutions that can still improve the best-found solution. Even though simple, we show that RAP expects to solve any discrete optimization problem that has n binary decision variables using at most $O(\mu\sqrt{2^n})$ operations, where μ represents the number of operations required to verify whether a solution is valid. To investigate the practical runtime of HBB and RAP, we use both procedures to solve 66000 instances of the binary knapsack problem (i.e., a subset of the aforementioned 108000 instances). The results show that both procedures are able to solve a binary knapsack problem that has n items using $O(n\sqrt{2^n})$ operations.

To further improve the performance of HBB and RAP, we assess the impact of several improvements. Because both procedures rely on GUM to find valid solutions, we first generalize GUM in the following way: (1)

if a valid solution was found assuming $\hat{m}$ solutions in the last search, we use $\hat{m}$ as the initial number of valid solutions in the next search (that considers a more restricted problem that has at most $\hat{m}$ valid solutions), (2) in each iteration, rather than letting $\hat{m} = \hat{m}/2$, we let $\hat{m} = \hat{m}/\kappa$, and (3) we stop the procedure after processing $\hat{m} = \tau$ rather than after $\hat{m} = 1$. The generalized procedure (GGUM) allows us to evaluate different strategies in which different samples of $\hat{m}$ are used to find a valid solution (note, however, that further research is required to develop a procedure that samples $\hat{m}$ such that the expected number of Grover iterations is minimized while maximizing the probability to find a valid solution; refer to the supplementary material for further ideas on how to improve sampling methods). Next, we also investigate the impact of using a heuristic to initialize the starting solution (rather than assuming an initial solution $V^* = 0$). To investigate the potential of techniques that improve the performance of the branch-and-bound of HBB, we also adopt an upper-bound procedure to fathom nodes (note that this upper-bound procedure cannot be used by RAP). The results of the experiment show that, even though we significantly improve performance, both procedures still require $O(n\sqrt{2^n})$ operations to solve a binary knapsack problem that has n items. The results, however, also show that HBB can outperform RAP if the structure of the problem can be exploited efficiently. In addition, the results indicate that HBB and RAP can be used as heuristics to find (near-) optimal solutions using (far) less than $O(n\sqrt{2^n})$ operations.

Next, we also generalize HBB and RAP, and show that they can be used to solve any discrete optimization problem using $O(\mu\sqrt{2^{nb}})$ operations, where 2^b is the number of discrete values that can be assigned to any of the n decision variables (see also Appendix D). As such, procedures HBB and RAP may be seen as general-purpose solvers that can solve well-known combinatorial problems as well as complex non-linear integer programming problems for which no dedicated algorithms exist. This is where the power of both procedures lies: their general applicability to solve any discrete optimization problem. Note that, even though HBB and RAP are general-purpose algorithms, they already match the performance of the best classical algorithm for solving the binary knapsack problem; a problem that has been studied for more than a century. Not only do HBB and RAP offer state-of-the-art performance when solving discrete optimization problems, they can also be used as heuristic procedures. In addition, as a branch-and-bound procedure, HBB allows to exploit the structure of a problem to further improve performance. In contrast, RAP can be used to solve problems that have no clear structure or that have a structure that is difficult to exploit. To combine the strengths of HBB and RAP, new procedures may be developed whose performance can be tailored depending on how effectively we can exploit the structure of the problem. The development of such new procedures, however, is left as a direction for future research (see also Appendix C).

Note that procedures HBB and RAP (or a combination thereof) can easily run in parallel. In the case of RAP, for instance, given i quantum processors, i instances of RAP can be running at the same time, and we expect an optimality gap of $1/(i+1)$ after completing a first iteration of RAP on all processors (e.g., if we have 20 quantum processors, we expect an optimality gap of less than 5 percent). In addition, we don't have to wait until completing an iteration of RAP to already share best-found solutions. In fact, for every value of $\hat{m}$ that is evaluated using GGUM, if a solution is found that improves the best-found solution, subsequent searches on all quantum processors can use this best solution as a lower bound. It would be interesting to investigate the computational gains that can be achieved this way. Similarly, the parallel execution of other Grover-based

algorithms may be an interesting topic for future research.

Our work also has several limitations that can be addressed. First, future research should evaluate the potential of procedures such as HBB and RAP to solve discrete optimization problems other than the binary knapsack problem. By comparing the performance of HBB, RAP, and state-of-the-art classical algorithms, we can identify problems that are more suited to be solved by quantum algorithms. Second, HBB and RAP are general-purpose algorithms that can be used to solve any discrete optimization problem. Therefore, procedures that are dedicated to solving a particular discrete optimization problem may outperform HBB, RAP, and perhaps even the best classical algorithms. The development of such dedicated procedures is a promising direction for future research. Third, procedures such as HBB and RAP can be extended such that they can also solve more general optimization problems (e.g., mixed-integer programming problems). Fourth, we focus on Grover-based algorithms. Next to Grover-based algorithms, however, other approaches (such as adiabatic quantum computation) may also be of interest. Fifth, HBB and RAP are exact procedures that, at best, can offer a quadratic speedup when compared to the classical algorithms. For large and/or complex problems, heuristic procedures are still required. Even though our results indicate that HBB and RAP can also be used as heuristic procedures, further research is required to investigate the potential of procedures such as HBB and RAP to quickly find good solutions. Last but not least, we assume the existence of a fault-tolerant universal quantum computer that can effectively run HBB and RAP. If we look at IBM's (ambitious) quantum roadmap, it becomes clear that it may take many more years before such a computer will hit the market. Once a fault-tolerant quantum computer does become available, however, state-of-the-art procedures such as HBB and RAP have the potential to revolutionize the field of discrete optimization.

The contributions of this paper may be summarized as follows: (1) we provide OR researchers with the essential tools to solve discrete optimization problems using Grover-based quantum algorithms, (2) we propose several quantum procedures and show how they can be used to solve the binary knapsack problem, (3) we demonstrate how to assess the performance of a quantum algorithm, (4) we show that our procedures match the best classical procedures when solving the binary knapsack problem, (5) we discuss several improvements that can further increase the performance of our procedures, (6) we illustrate that our procedures can also be used as efficient heuristics to find (near-) optimal solutions, (7) we show that our procedures can be generalized such that they can be used as general-purpose solvers that have the potential to revolutionize the field of discrete optimization, and (8) we provide a myriad of directions for future research.

References

Aaronson, S. 2018. *Introduction to Quantum Information Science*. Lecture Notes, UT Austin.

Albash, T., Lidar, D.A. 2018. Adiabatic quantum computation. *Rev. Mod. Phys.*, **90**(1), 015002.

Ambainis, A., Balodis, K., Iraids, J., Kokainis, M., Prūsis, K., and Vihrovs J. 2018. Quantum speedups for exponential-time dynamic programming algorithms, unpublished preprint at https://doi.org/10.48550/arXiv.1807.05209.

Babbush, R., McClean, J.R., Newman, M., Gidney, C., Boixo, S., and Neven, H. 2021. Focus beyond quadratic speedups for error-corrected quantum advantage. *PRX quantum*, **2**(1), 010103.

Benioff, P. 1980. The computer as a physical system: A microscopic quantum mechanical Hamiltonian model of computers as represented by Turing machines. *J. Stat. Phys.*, **22**(5), 563–591.

Bennett, C. H., Bernstein, E., Brassard, G., and Vazirani, U. 1997. Strengths and weaknesses of quantum computing. *SIAM J. Comput.*, **26**(5), 1510-1523.

Blekos, K., Brand, D., Ceschini, A., Chou, C.-H., Li, R.-H., Pandya, K., and Summer, A. A Review on Quantum Approximate Optimization Algorithm and its Variants, unpublished preprint at https://doi.org/10.48550/arXiv.2306.09198.

Born, M. 1926. Zur Quantenmechanik der Stoßvorgänge. *Z. Phys.*, **37**(12), 863–867.

Boyer, M., Brassard, G., Høyer, P., and Tapp, A. 1996. *Tight Bounds on Quantum Searching*. Proceedings of the Fourth Workshop on Physics and Computation, New England Complex Systems Institute.

Campbell, E., Khurana, A., and Montanaro, A. 2019. Applying quantum algorithms to constraint satisfaction problems. *Quantum.*, **3**, 167.

Cheng, B., Deng, X.H., Gu, X., He, Y., Hu, G., Huang, P., Li, J., Lin, B.C., Lu, D., Lu, Y., Qiu, C., Wang, H., Xin, T., Yu, S., Yung, M.H., Zeng, J., Zhang, S., Zhong, Y., Peng, X., Nori, F., and Yu, D. 2023. Noisy intermediate-scale quantum computers. *Front. Phys.*, **18**.

Cleve, R., Ekert, A., Macchiavello, C., and Mosca, M. 1998. Quantum algorithms revisited. *Proc. R. Soc. Lond. A*, **454**(1969), 339–354.

Creemers, S., Pérez, L.F. 2023. Discrete optimization: Limitations of existing quantum algorithms, unpublished preprint at https://dx.doi.org/10.2139/ssrn.4527268.

Dantzig, G.B. 1957. Discrete-variable extremum problems. *Operations Research*, **5**(2), 266–288.

Deutsch, D. 1985. Quantum theory, the Church–Turing principle and the universal quantum computer. *Proc. R. Soc. Lond. A*, **400**(1818), 97–117.

Deutsch, D., Jozsa, R. 1992. Rapid solution of problems by quantum computation. *Proc. R. Soc. Lond. A*, **439**(1907), 553—558.

Dirac, P. A. M. 1939. A new notation for quantum mechanics. *Math. Proc. Camb. Philos. Soc.*, **35**(3), 416–418.

Farhi, E., Goldstone, J., and Gutmann, S. A quantum approximate optimization algorithm, unpublished preprint at https://doi.org/10.48550/arXiv.1411.4028.

Feynman, R. P. 1982. Simulating physics with computers. *Int. J. Theor. Phys.*, **21**(6-7), 467–488 .

Glover, F., Kochenberger, G., and Hennig, R. 2022. Quantum bridge analytics I: a tutorial on formulating and using QUBO models. *Ann Oper Res*, **314**(1), 141–183.

Grover, L. K. 1996. A fast quantum mechanical algorithm for database search. *Proc. Annu. ACM Symp. Theory Comput.*, 212–219.

Hidary, J.D. 2021. *Quantum Computing: An Applied Approach*. Springer Nature Switzerland AG, Cham, Switzerland.

Kellerer, H., Pferschy, U., and Pisinger, D. 2004. *Knapsack Problems*. Springer Berlin Heidelberg, Berlin, Heidelberg.

Kurowski, K., Pecyna, T., Slysz, M., Różycki, R., Waligóra, G., and Węglarz, J. 2023. Application of quantum approximate optimization algorithm to job shop scheduling problem. *European Journal of Operational Research*, **310**(2), 518–528.

MacQuarrie, E.R., Simon, C., Simmons, S., and Maine, E. 2020. The emerging commercial landscape of quantum computing. *Nature Reviews Physics*, **2**, 596–598.

McKinsey 2021. *Quantum computing: An emerging ecosystem and industry use cases*. McKinsey.

McKinsey 2023. *Quantum technology monitor*. McKinsey.

Montanaro, A. 2016. Quantum walk speedup of backtracking algorithms, unpublished preprint at https://doi.org/10.48550/arXiv.1509.02374.

Nannicini, G. 2020. An introduction to quantum computing, without the physics. *SIAM Review*, **62**(4), 936–981.

Nielsen, M., Chuang, I. 2010. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, Cambridge.

Parker, R.G., Rardin, R.L. 1988. *Discrete Optimization*. Academic Press, Inc., San Diego.

Pisinger, D. 2005. Where are the hard knapsack problems? *Comput. Oper. Res.*, **32**(9), 2271–2284.

Preskill, J. 2018. Quantum Computing in the NISQ era and beyond. *Quantum 2*, **2**, 79–99.

Shor, P.W. 1994. Algorithms for quantum computation: Discrete logarithms and factoring. *Proc. Annu. Symp. FOCS*, 124–134.

Wolsey, L.A. 2021. *Integer Programming*. Wiley, Hoboken.

Appendix

A. Probability to measure target basis state after I Grover iterations

Let $P(n, m, I)$ denote the probability to measure one of the m target basis state after performing I Grover iterations in a system that has 2^n basis states. To obtain $P(n, m, I)$, we use Algorithm 7. In a first step of Algorithm 7, we initialize the current probability amplitude of the other basis states (α_o), the current probability amplitude of the target basis states (α_t), and the current iteration (i). Next, in each iteration i, we calculate the average probability amplitude ($\bar{\alpha}$) as the weighted sum of the probability amplitudes of the target and the other basis states. We use the average probability amplitude to update the probability amplitude of the target and other states (i.e., we reflect α_t and α_o about $\bar{\alpha}$) and return $P(n, m, I)$ as α_t^2 if $i \geq I$.

Algorithm 7: Procedure to determine probability to measure a target basis state in a system that has n qubits and m target basis states.

initialize $\alpha_o = (\sqrt{2^n})^{-1}$, $\alpha_t = -\alpha_o$, and $i = 1$;
while $i < I$ **do**
  $\bar{\alpha} = 2^{-n}((2^n - m)\alpha_o + m\alpha_t)$;
  $\alpha_t = \alpha_t - 2\bar{\alpha}$;
  $\alpha_o = 2\bar{\alpha} - \alpha_o$;
  $i = i + 1$;
return α_t^2;

B. Assessing performance of a quantum algorithm

The most straightforward way to assess the performance of a quantum algorithm is to implement the algorithm on a quantum computer (or a simulation thereof). Current quantum computers and simulators, however, are limited to solving trivial optimization problems. In addition, access to current quantum computers is limited and/or expensive. Therefore, in this paper, we adopt a different approach (for a tutorial on how to implement quantum algorithms on existing quantum computers/simulators, refer to, for example, the documentation of IBM's Qiskit or Google's Cirq). Rather than implementing the full circuit of a quantum algorithm, we calculate the probability of each possible outcome of the subcircuits (i.e., subroutines) used by the algorithm. Using these probabilities, we can model the transitions between different system states that result from applying the algorithm. In our case, rather than implementing Grover's algorithm, we use Algorithm 7 to obtain $P(n, m, I)$; the probability of measuring one of the m valid solutions in a system that has 2^n valid solutions after performing I Grover iterations. Using $P(n, m, I)$, we can determine the probability to transition from one system state to another. We propose two methods to assess the performance of a quantum algorithm: the exact evaluation of the performance using a Discrete-Time Markov Chain (DTMC) and the simulation of this DTMC using Monte Carlo simulation. In what follows, we use both methods to assess the performance of `BSP`.

A DTMC can be used to model the evolution of a quantum algorithm starting from an initial state (i.e., the start of the algorithm) until we end up in an absorbing state (i.e., the completion of the algorithm). In case of `BSP`, a state may be defined by tuple $(V^*, V_{\min}, V_{\max}, \widehat{m}, \text{GI})$, where GI represents the number of Grover iterations that have been performed to reach state $(V^*, V_{\min}, V_{\max}, \widehat{m}, \text{GI})$. When departing from state $(V^*, V_{\min}, V_{\max}, \widehat{m}, \text{GI})$, the possible transitions and their probabilities are listed in Table 7. If $V_{\min} > V_{\max}$, the algorithm stops and we end up in an absorbing state. If, on the other hand, $V_{\min} \leq V_{\max}$, there are three possibilities. If a valid solution is found, we transition to a state where V^* and $V_{\min}$ are updated. If no valid solution is found, we either transition to a state where $\widehat{m}$ is updated or where $V_{\max}$ and $\widehat{m}$ are updated (if $\widehat{m} = 1$). The probability to find a valid solution in state $(V^*, V_{\min}, V_{\max}, \widehat{m}, \text{GI})$ is given by $P(n, m_V, I(n, \widehat{m}))$, where m_V denotes the number of valid solutions that have a value of at least $V = \lfloor 0.5(V_{\min} + V_{\max}) \rfloor$.

Table 7. Possible DTMC transitions when departing from state $(V^*, V_{\min}, V_{\max}, \widehat{m}, \text{GI})$.

resulting state	transition probability	conditions
$(V^*, V_{\min}, V_{\max}, \widehat{m}, \text{GI})$	1	$V_{\min} > V_{max}$
$(V, (V+1), V_{\max}, 2^n, (\text{GI} + I(n, \widehat{m})))$	$P(n, m_V, I(n, \widehat{m}))$	if $V_{\min} \leq V_{\max}$
$(V^*, V_{\min}, V_{\max}, \widehat{m}/2, (\text{GI} + I(n, \widehat{m})))$	$1 - P(n, m_V, I(n, \widehat{m}))$	$V_{\min} \leq V_{\max} \wedge \widehat{m} > 1$
$(V^*, V_{\min}, (V-1), 2^n, (\text{GI} + I(n, \widehat{m})))$	$1 - P(n, m_V, I(n, \widehat{m}))$	if $V_{\min} \leq V_{\max} \wedge \widehat{m} = 1$

We now illustrate how to use a DTMC to assess the performance of `BSP` when solving the example binary knapsack problem. Table 8 lists the states, the probabilities to visit each of the states, and the transitions that are possible from each state. For instance, we have a 4.785 percent probability to visit state 14 (with $V^* = 3$, $V_{\min} = 4$, $V_{\max} = 6$, $\widehat{m} = 2$, and GI $= 3$). In state 14, $V = 5$ and there is only one valid solution (i.e., $m_V = 1$). As a result, if we perform $I(n, \widehat{m}) = I(3, 2) = 2$ Grover iterations, we have a 94.53 percent

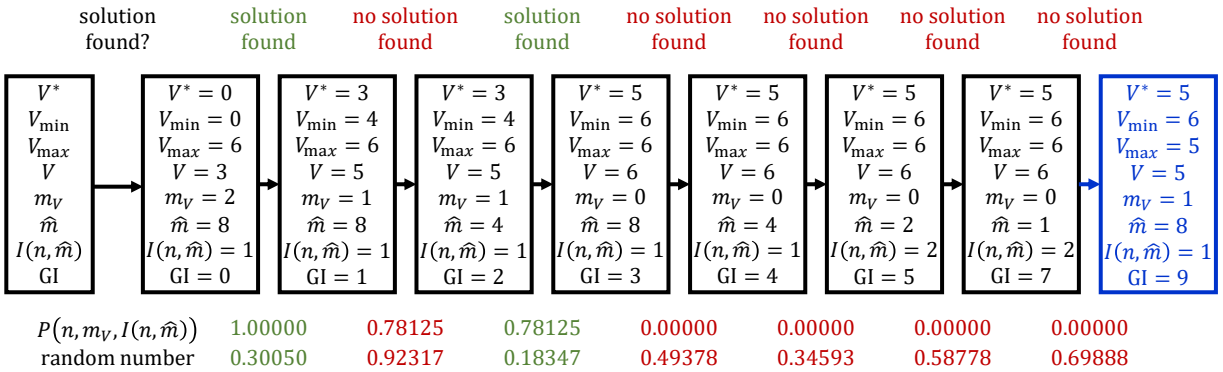
probability to measure a valid solution and to transition to state 15 (with updated $V^* = 5$ and $V_{\min} = 6$). There is, however, also a 5.469 percent probability that we transition to state 20 (with updated $\hat{m} = 1$). By observing the probabilities to end up in each of the absorbing states, we can determine: (1) the best-case number of required Grover iterations, (2) the worst-case number of required Grover iterations, and (3) the expected number of required Grover iterations. For the example binary knapsack project, we require 8, 13, and 8.319 Grover iterations, respectively. The probabilities to end up in each of the absorbing states can also be used to verify the probability to identify an optimal solution (which amounts to 99.9857 percent for the example binary knapsack problem).

Table 8. DTMC that assesses the performance of BSP when solving the example binary knapsack problem. For each state of the DTMC we list the visit probability, the tuple, the transition if a solution is found, and the transition if no solution is found. The initial state is indicated in gray and absorbing states are indicated in blue. Probabilities are expressed as percentages.

state	visit prob.	departure state tuple					target state	trans. prob.	target state	trans. prob.
		V^*	$V_{\min}$	$V_{\max}$	$\hat{m}$	GI				
1	100	0	0	6	8	0	2	100	NA	0
2	100	3	4	6	8	1	3	78.13	8	21.88
3	78.125	5	6	6	8	2	NA	0	4	100
4	78.125	5	6	6	4	3	NA	0	5	100
5	78.125	5	6	6	2	4	NA	0	6	100
6	78.125	5	6	6	1	6	NA	0	7	100
7	78.125	5	6	5	8	8	7	100	NA	0
8	21.875	3	4	6	4	2	9	78.13	14	21.88
9	17.08984	5	6	6	8	3	NA	0	10	100
10	17.08984	5	6	6	4	4	NA	0	11	100
11	17.08984	5	6	6	2	5	NA	0	12	100
12	17.08984	5	6	6	1	7	NA	0	13	100
13	17.08984	5	6	5	8	9	13	100	NA	0
14	4.785156	3	4	6	2	3	15	94.53	20	5.469
15	4.523468	5	6	6	8	5	NA	0	16	100
16	4.523468	5	6	6	4	6	NA	0	17	100
17	4.523468	5	6	6	2	7	NA	0	18	100
18	4.523468	5	6	6	1	9	NA	0	19	100
19	4.523468	5	6	5	8	11	19	100	NA	0
20	0.261688	3	4	6	1	5	21	94.53	26	5.469
21	0.247377	5	6	6	8	7	NA	0	22	100
22	0.247377	5	6	6	4	8	NA	0	23	100
23	0.247377	5	6	6	2	9	NA	0	24	100
24	0.247377	3	4	4	1	11	NA	0	25	100
25	0.247377	5	6	5	8	13	25	100	NA	0
26	0.014311	3	4	4	8	7	27	78.13	28	21.88
27	0.011181	4	5	4	8	8	27	100	NA	0
28	0.003131	3	4	4	4	8	29	78.13	30	21.88
29	0.002446	4	5	4	8	9	29	100	NA	0
30	0.000685	3	4	4	2	9	31	94.53	32	5.469
31	0.000647	4	5	4	8	11	31	100	NA	0
32	0.000037	3	4	4	1	11	33	94.53	34	5.469
33	0.000035	4	5	4	8	13	33	100	NA	0
34	0.000002	3	4	3	8	13	34	100	NA	0

Next to using an exact DTMC approach, we can also use Monte Carlo simulation to approximate the performance of a quantum algorithm. In this case, random numbers are drawn to determine the transitions

Figure 16. Example simulation iteration when using Monte Carlo simulation to assess the performance of BSP when solving the example binary knapsack problem. For each run of Grover's algorithm, we compare the probability of measuring a valid solution (i.e., $P(n, m, I(n, \hat{m}))$) and a random number. If the random number if smaller-than-or-equal-to $P(n, m, I(n, \hat{m}))$, a valid solution is found (indicated in green), otherwise no solution is found (indicated in red). The simulation completes if $V_{\min} > V_{\max}$ (indicated in blue).



between system states of the DTMC. A simulation iteration starts in an initial state (that corresponds to the start of the algorithm) and ends in an absorbing state (that corresponds to the completion of the algorithm). After performing sufficient simulation iterations, we can approximate the probability to end up in each of the absorbing states. In turn, these probabilities allow us to approximate the expected performance of the quantum algorithm as well as the probability that an optimal solution is identified. In case of BSP, the state of the system may once again be defined by tuple $(V^*, V_{\min}, V_{\max}, \hat{m}, \text{GI})$ and all possible transitions are listed in Table 7. In any simulation iteration, if $V_{\min} > V_{\max}$, we reached an absorbing state and the simulation iteration is finished. If, on the other hand, $V_{\min} \leq V_{\max}$, we draw a random number and verify whether this random number is larger than $P(n, m_V, I(n, \hat{m}))$. If the random number is larger, no valid solution was found, and a transition is made towards a state where we either update $\hat{m}$ and/or $V_{\max}$. Otherwise, if the random number is smaller-than-or-equal-to $P(n, m_V, I(n, \hat{m}))$, a valid solution was found, and we transition towards a state where we update V^* and $V_{\min}$. Note that, if a valid solution is found in state $(V^*, V_{\min}, V_{\max}, \hat{m}, \text{GI})$ (with probability $P(n, m_V, I(n, \hat{m}))$), we can obtain this solution by randomly drawing a solution from the set of m_V valid solutions (each valid solution has probability m_V^{-1} of being measured).

Next, we illustrate how Monte Carlo simulation can be used to approximate the performance of BSP when solving the example binary knapsack problem. Figure 16 provides an example of a single simulation iteration. In each simulation iteration, we start from state $(V^*, V_{\min}, V_{\max}, \hat{m}, \text{GI}) = (0, 0, 6, 8, 0)$. In this state, $V = 3$, $m_V = 2$, and we are certain to identify a valid solution after a single iteration of Grover's algorithm (i.e., $P(n, m_V, I(n, \hat{m})) = P(3, 2, 1) = 1$). As a result, regardless of the random number that is drawn, we transition to state $(3, 4, 6, 8, 1)$. In this state, $V = 5$, $m_V = 1$, and there is only a 78.125 percent chance that we identify a valid solution. There are two options. First, if a random number is drawn that is larger than 0.78125, we say that no solution was found and transition to state $(3, 4, 6, 4, 2)$. If, on the other hand, we draw a random number smaller-than-or-equal-to 0.78125, we say that a solution was found and we transition to state $(5, 6, 6, 8, 2)$. This process continues until we end up in an absorbing state (state $(5, 6, 6, 8, 1, 9)$ in Figure 16). After performing a sufficient number of iterations, we can once again approximate: (1) the

best-case number of Grover iterations, (2) the worst-case number of Grover iterations, and (3) the expected number of Grover iterations. In addition, we can also approximate the probability to identify an optimal solution by keeping track of how often an optimal solution value was observed at the end of a simulation iteration.

Both the DTMC and the Monte Carlo simulation method allow to analyze the performance of a quantum algorithm. Whereas Monte Carlo simulation provides an approximation of the performance, a DTMC can be used to obtain exact results. Both methods, however, have one drawback: we have to be able to determine the probabilities to transition from one system state to another. In the case of Grover-based algorithms (i.e., in our case), this implies that we need to determine $P(n, m_V, I(n, \widehat{m}))$. In order to determine $P(n, m_V, I(n, \widehat{m}))$, however, we need to know m_V; we need to know the number of valid solutions that have a value of at least V. In general, we do not know m_V. In this paper, we use a brute-force approach to calculate m_V for each possible value of V. With 2^n possible solutions, this is still feasible for n up to 30. For larger n, however, a brute-force approach would no longer be feasible (the use of memory and/or computation times becomes prohibitive). Therefore, for larger n, approximation algorithms may be used to approximate m_V. For instance, Monte Carlo simulation can be used to draw a sample of random solutions that can be used to approximate the distribution of m_V for different values of V. The larger the sample, the more accurate our estimate of m_V will be. The investigation of such approximations, as well as the study of other methods to assess the performance of quantum algorithms, is left as a direction for future research.

C. Combining procedures HBB and RAP

In this section, we combine both HBB and RAP. Whereas HBB has the disadvantage that a significant amount of time is spent on proving that a (partial) solution is optimal, it has the advantage that it can exploit the structure of a problem (resulting in the removal of subsets of solutions from the search space). RAP, on the other hand, minimizes the time spent on proving that a solution is optimal, however, it always searches the entire solution space. To combine the strengths of both procedures, HBB can be equipped with RAP at deeper levels of the search tree. This way: (1) we benefit from the speedup created by the branch-and-bound procedure at higher levels of the search tree, (2) we spend less time to prove that a partial solution is optimal in deeper levels of the tree, and (3) we also counter the drawback of RAP because it is used to find a partial solution in only a subset of the solution space. In addition, note that at higher levels of the search tree, there is not so much difference between the searches of HBB and RAP (i.e., at higher levels of the tree, both HBB and RAP search the same solution space). As a result, it makes sense to use RAP mainly at deeper levels of the search tree.

Algorithm 8 (hereafter referred to as RAHBB) presents one possible implementation that uses RAP if the number of Grover iterations required by HBB exceeds a given threshold nv. Procedure RAHBB also uses upper bound UBi to fathom nodes and is equipped with GGUM rather than with GUM. We use RAHBB to solve the 66000 instances of the binary knapsack problem assuming the same settings as those used in the "full case" discussed in the Section 9 (i.e., we use a lower bound heuristic to set an initial solution, we let $\tau = 4$, we use upper-bound procedure UBi to fathom nodes, and we use the number of valid solutions of the last search as the initial number of valid solutions in the subsequent search). The results are summarized in

Figure 17. Figure 17 presents the number of operations (divided by $n\sqrt{2^n}$) that are required by RAHBB for various values of υ. Figure 17 also reports the performance of RAP (top left panel) and HBB (bottom right panel). We can observe that, for sufficiently small υ, the performance of RAHBB is almost identical to that of RAP (i.e., RAHBB skips HBB, and immediately performs RAP). On the other hand, if υ is sufficiently large, RAHBB is identical to HBB (i.e., RAHBB never performs RAP). Figure 17 shows that, depending on our ability to exploit the structure of the problem, a combination of HBB and RAP may be used to reduce the overall expected number of operations.

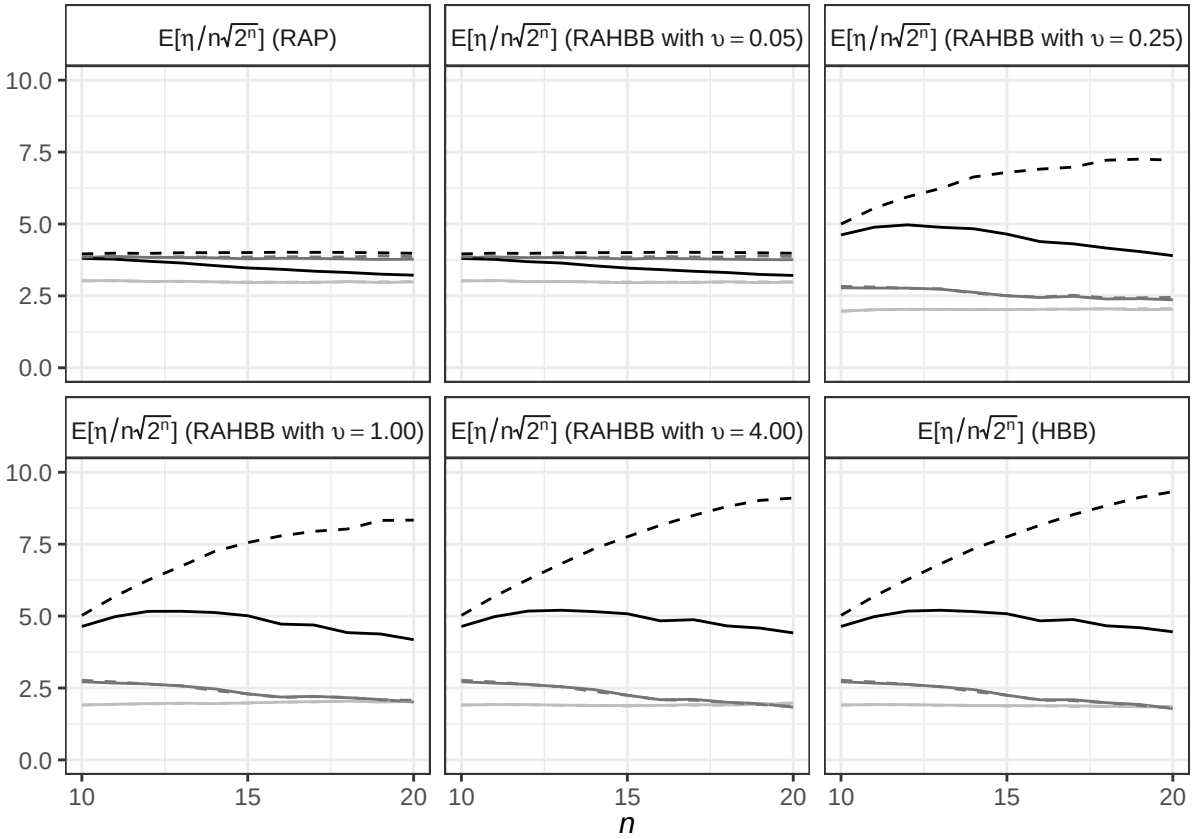
Algorithm 8: Hybrid branch-and-bound procedure that uses a random-ascent procedure if the number of required Grover iterations exceeds $n\upsilon$. Both procedures use GGUM to measure a valid solution.

initialize $\mathbf{x}^*$ and V^* using lower bound heuristic and start procedure RAHBB$(\emptyset, 0, W, 2^n, 1)$;
Procedure RAHBB$(\mathbf{x}, V, R, \widehat{m}, i)$
  if UBi$(V, R, i) > V^*$ **then**
    if $n - i < \omega$ **then**
      use classical algorithm to find $\mathbf{x}_{in}^\star \in \Omega_{in}$ that has highest value $V_{in}^\star > V^* - V$ and $W_{in}^\star \leq R$;
      if $\mathbf{x}_{in}^\star \neq \emptyset$ **then**
        update best-found solution $\mathbf{x}^* = \mathbf{x}_{1(i-1)} \cup \mathbf{x}_{in}^\star$ and its value $V^* = V + V_{in}^\star$;
    else if $I(n - i, \widehat{m}) \leq n\upsilon$ **then**
      if $w_i \leq R$ **then**
        Let $x_i = 1$, $\mathbf{x}_{1i} = \mathbf{x}_{1(i-1)} \cup \{x_i\}$, $V' = V + v_i$, and $R' = R - w_i$;
        Use GGUM$(n - i, \widehat{m})$ to obtain $\widehat{m}'$ and to measure $\mathbf{x}_{(i+1)n} \in \Omega_{(i+1)n}$ that has value $V_{(i+1)n} > V^* - V'$ and $W_{(i+1)n} \leq R'$;
        if $\mathbf{x}_{(i+1)n} \neq \emptyset$ **then**
          update best-found solution $\mathbf{x}^* = \mathbf{x}_{1i} \cup \mathbf{x}_{(i+1)n}$ and its value $V^* = V' + V_{(i+1)n}$;
          RAHBB$(\mathbf{x}_{1i}, V', R', \widehat{m}', i + 1)$;
      Let $x_i = 0$, $\mathbf{x}_{1i} = \mathbf{x}_{1(i-1)} \cup \{x_i\}$, and run RAHBB$(\mathbf{x}, V, R, \widehat{m}, i + 1)$;
    else
      do
        Use GGUM$(n - i, \widehat{m})$ to obtain $\widehat{m}'$ and to measure $\mathbf{x}_{(i+1)n} \in \Omega_{(i+1)n}$ that has value $V_{(i+1)n} > V^* - V'$ and $W_{(i+1)n} \leq R'$;
        if $\mathbf{x}_{(i+1)n} \neq \emptyset$ **then**
          Let $\widehat{m} = \widehat{m}'$ and update best-found solution $\mathbf{x}^* = \mathbf{x}_{1i} \cup \mathbf{x}_{(i+1)n}$ and its value $V^* = V' + V_{(i+1)n}$;
      while $\mathbf{x} \neq \emptyset$;

D. Generalizing procedures HBB and RAP

Until now, we have considered discrete optimization problems with binary decision variables. Procedures HBB and RAP (and by extension RAHBB), however, can easily be adapted to solve discrete optimization problems that consider discrete decision variables. If we assume that a decision variable can take on 2^b values,

Figure 17. Expected number of operations divided by $n\sqrt{2^n}$ required to solve a binary knapsack problem using `HBB`, `RAP`, and `RAHBB` for various values of υ, when solving 1000 binary knapsack problems for each problem set, for each value of n : $10 \leq n \leq 20$ (lines of sets 1 and 4 are light gray, sets 2 and 5 are dark gray, sets 3 and 6 are black, sets 1, 2, and 3 are full, and sets 4, 5, 6 are dashed).



we require b qubits to represent the decision variable (i.e., in the case of binary decision variables, $b = 1$, and a single qubit suffices to represent a decision variable). If each of the n decision variables requires b qubits, `GUM` is applied to nb qubits rather than to n qubits. As a result, there are 2^{nb} possible solutions, and `GUM` will evaluate up to $\log_2(2^{nb}) = nb$ values of $\widehat{m}$. For each value of $\widehat{m}$, `GUM` performs $I(2^{nb}, \widehat{m})$ Grover iterations (that each require two calls to function $f_{\mathbf{x}}$) and a single call to function $f_{\mathbf{x}}$ to verify the validity of the measured solution. If we assume it takes μ operations to evaluate a solution, `HBB` and `RAP` (or a combination thereof) require $O(\mu\sqrt{2^{nb}})$ operations.

Supplementary material for: Discrete optimization: A quantum revolution?

Stefan Creemers*
IESEG School of Management, Lille, France, s.creemers@ieseg.fr
KU Leuven, Leuven, Belgium, stefan.creemers@kuleuven.be

Pérez Luis Fernando
IESEG School of Management, Lille, France, l.perezarmas@ieseg.fr

*Corresponding author

A. Universal Quantum Computing

Similar to classical computing, quantum computing uses logic gates to perform operations. In contrast to the gates of classical computing, however, quantum gates have to be reversible, and can be represented by unitary matrices (i.e., a matrix U is unitary if $UU^{-1} = UU^{\mathsf{T}} = I$, where I is the identity matrix). In what follows, we first introduce the X, CX, and Toffoli quantum gates, and show how these gates can be used to create the quantum equivalent of the AND, OR, XOR, and NOT gates of classical computing (note, however, that there are many other quantum gates; refer to, for example, Aaronson (2018) for an overview of other quantum gates).

The X gate is the quantum equivalent of the classical NOT gate and transforms input state $|0\rangle$ into output state $|1\rangle$ (and vice versa). Formally, X is a two-by-two matrix that operates as follows:

$$\mathrm{X}\,|0\rangle = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} = |1\rangle\,, \quad \mathrm{X}\,|1\rangle = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = |0\rangle\,.$$

The CX (or controlled-X) gate has two input qubits: a target qubit on which an X operation may be performed, and a control qubit that determines whether or not the X operation is performed. If the control qubit is in basis state $|1\rangle$, the X operation is performed on the target qubit. Otherwise, if the control qubit is in basis state $|0\rangle$, the target qubit remains unchanged. Formally, CX is a four-by-four matrix that operates as follows:

$$\mathrm{CX}\,|00\rangle = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = |00\rangle\,.$$

One can verify that $\mathrm{CX}\,|10\rangle = |11\rangle$, $\mathrm{CX}\,|01\rangle = |01\rangle$, and $\mathrm{CX}\,|11\rangle = |10\rangle$.

The Toffoli gate is a controlled-CX (CCX) gate that has two control qubits. Only if both control qubits are in basis state $|1\rangle$, the X operation is performed on the target qubit. Formally, CCX is an eight-by-eight matrix that operates as follows:

$$\mathrm{CCX}\,|000\rangle = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} = |000\rangle\,.$$

One can verify that $\mathrm{CCX}\,|100\rangle = |100\rangle$, $\mathrm{CCX}\,|010\rangle = |010\rangle$, $\mathrm{CCX}\,|110\rangle = |111\rangle$, $\mathrm{CCX}\,|001\rangle = |001\rangle$, $\mathrm{CCX}\,|101\rangle = |101\rangle$, $\mathrm{CCX}\,|011\rangle = |011\rangle$, and $\mathrm{CCX}\,|111\rangle = |110\rangle$.

We can use the X, CX, and CCX gates to create circuits that mimic the classical AND, OR, XOR, and NOT gates: the quantum equivalent of the AND gate corresponds to a CCX gate with target qubit initialized as $|0\rangle$, the quantum equivalent of the OR gate is a series of a CX, X, CCX, and X gate with target qubit initialized as $|0\rangle$, the quantum equivalent of the XOR gate is a series of two CX gates with target qubit initialized as $|0\rangle$, and the quantum equivalent of the NOT gate is simply the X gate. The circuits of all aforementioned gates are shown in Figure 1.

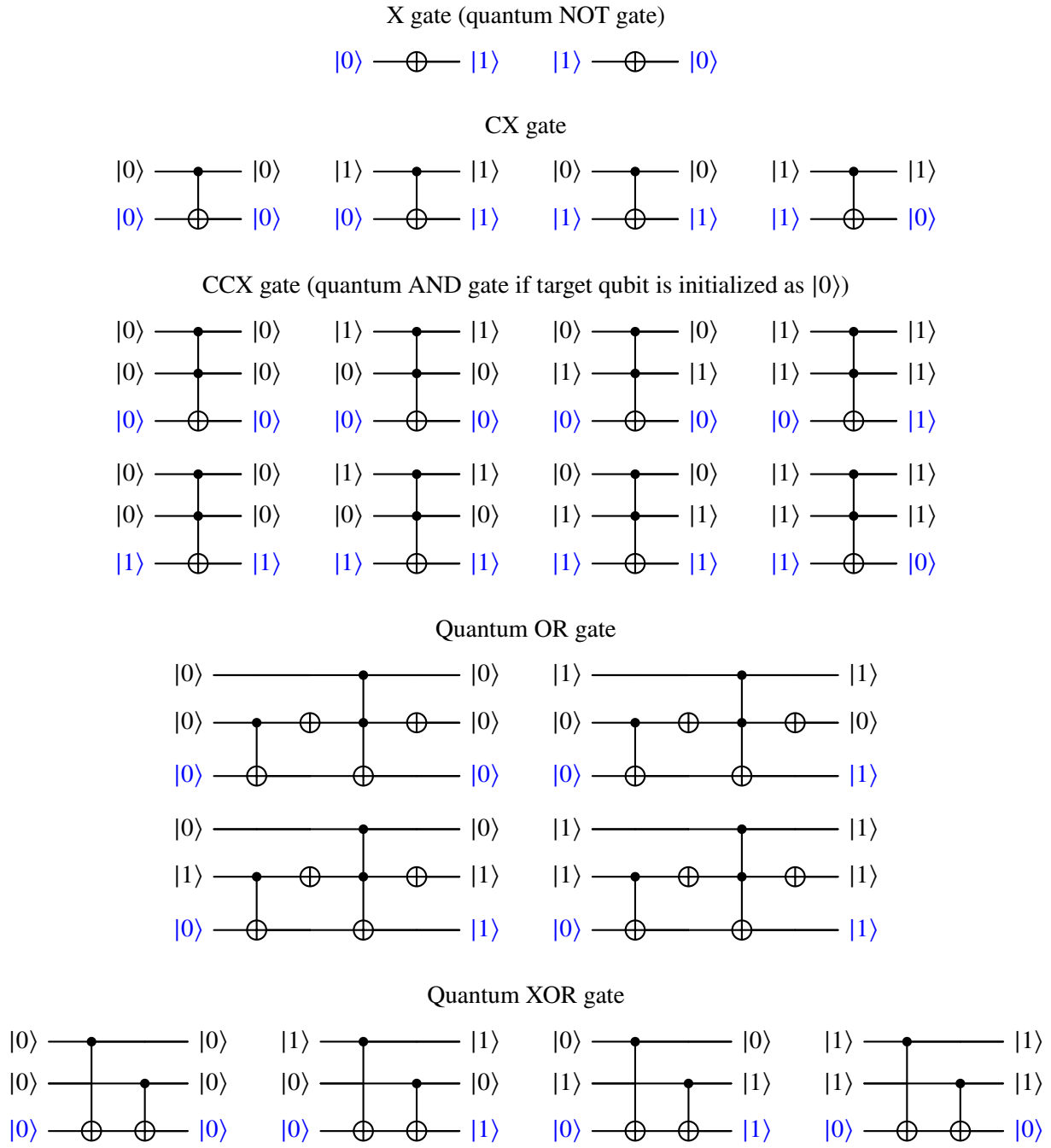


Figure 1. Circuits of the X, CX, and CCX quantum gates, as well as the quantum equivalents of the classical AND, OR, XOR, and NOT gates. The target qubit is indicated in blue.

The NOT, AND, and OR operation suffice to establish universal classical computing. As a result, the X, CX, and CCX gate (i.e., the quantum equivalent of the NOT, AND, and OR operation on a classical computer) allow to perform any classical operation on a quantum computer (see also Deutsch (1989) who has shown that any classical operation can be implemented on a quantum computer using only a polynomial overhead). Quantum computers, however, are not limited to classical operations. To establish universal quantum computing, we are looking for a set of quantum gates that allow us to approximate any quantum gate (i.e., any unitary operator) efficiently within a given error bound (following the Solovay-Kitaev theorem). Several sets of universal quantum gates exist. For instance, in 2003, Aharonov has shown that the CCX and the Hadamard gate form a set of universal quantum gates.

B. Limitations of the $I(n, m)$ approximation

To maximize the probability to find one of the m target basis states in a set of 2^n basis states, we perform $I(n, m) = \left\lfloor \pi 4^{-1} \sqrt{m^{-1} 2^n} \right\rfloor$ Grover iterations. In practice, however, we may need far more than $I(n, m)$ Grover iterations. For instance, if $m > {}^{2^n}/_2$, it is well known that approximation $I(n, m)$ may result in a low probability to measure a target basis state (e.g., in a system that has $n = 3$ qubits and $m = 6$ target basis states, $I(3, 6) = 1$, and the probability to measure a target state after 1 Grover iteration equals 0). To resolve this problem, it has been suggested to add a "dummy" qubit. By adding a dummy qubit, the number of basis states effectively doubles while the number of target basis states remains the same. As a result, the number of target basis states is always smaller than (or equal to) half of the number of basis states. Doubling the number of basis states, however, also doubles the search space. As such, this is not an ideal (nor elegant) solution. In addition, even if $m \leq {}^{2^n}/_2$, approximation $I(n, m)$ may result in a low probability to measure a target basis state. For instance, consider a system that has $n = 6$ qubits that has $m = 31$ target basis states (and 33 other basis states). In such a system, we only have a 54.68 percent probability to measure a target basis state after $I(6, 31) = 1$ Grover iteration. In order to be at least 95 percent certain to measure a target basis state, we need 18 Grover iterations. This is also illustrated in Figure 2 that shows the probability to measure one of the 31 target basis states after each Grover iteration. Unfortunately, the number of required Grover iterations is rather sensitive to the number of target basis states. For instance, in the example of 6 qubits, if we have 30, 29, 28, or 27 target basis states, we need 10, 7, 5, and 5 Grover iterations to be at least 95 percent certain to measure any of the target basis states, respectively. For additional examples, refer to Figure 3 that shows the probability to measure any of the $m : m \in \{17, 22, 28, 29, 30, 31\}$ target basis states after each Grover iteration in a system that has $n = 6$ qubits. Figure 3 also illustrates why a sine function should not be used to approximate $P(n, m, I)$ for all values of m (a sine function can only be used to approximate the probability to measure any of the m target basis states if m is sufficiently small; see, for example, Figure 6 in the main paper). Investigating (sine) approximations of $P(n, m, I)$ is left as a direction for future research.

In order to determine the number of iterations required by Grover's algorithm, we use Algorithm 9 (hereafter referred to as procedure $\texttt{GI}\rho$). For a system of n qubits and m target basis states, $\texttt{GI}\rho$ returns the minimum number of iterations required to be at least ρ percent certain to measure a target basis state. In a first step of the procedure, we initialize the current probability amplitude of the other basis states (α_o), the current

Figure 2. Probability of measuring a target basis state after each Grover iteration for a system that has 6 qubits and 31 target basis states (for reference, probability 0.95 is indicated by the blue dashed line).

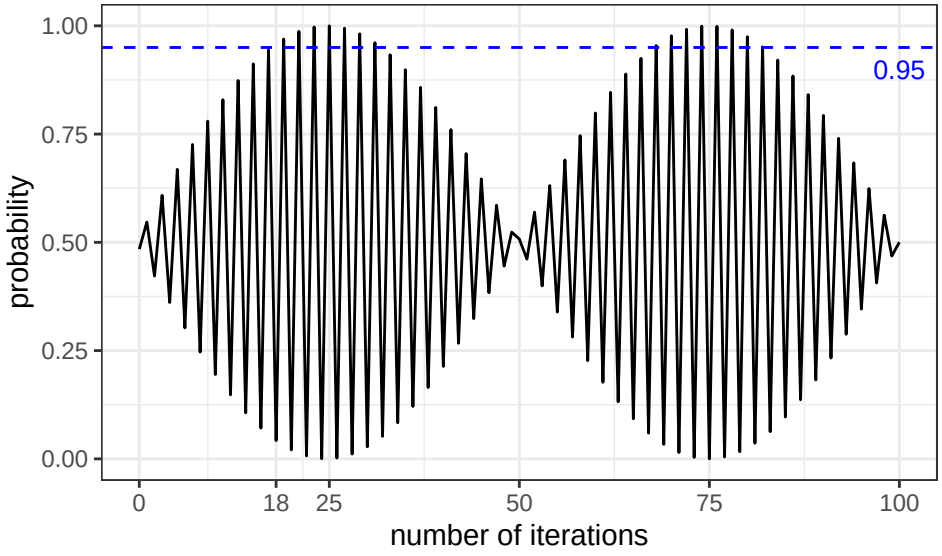
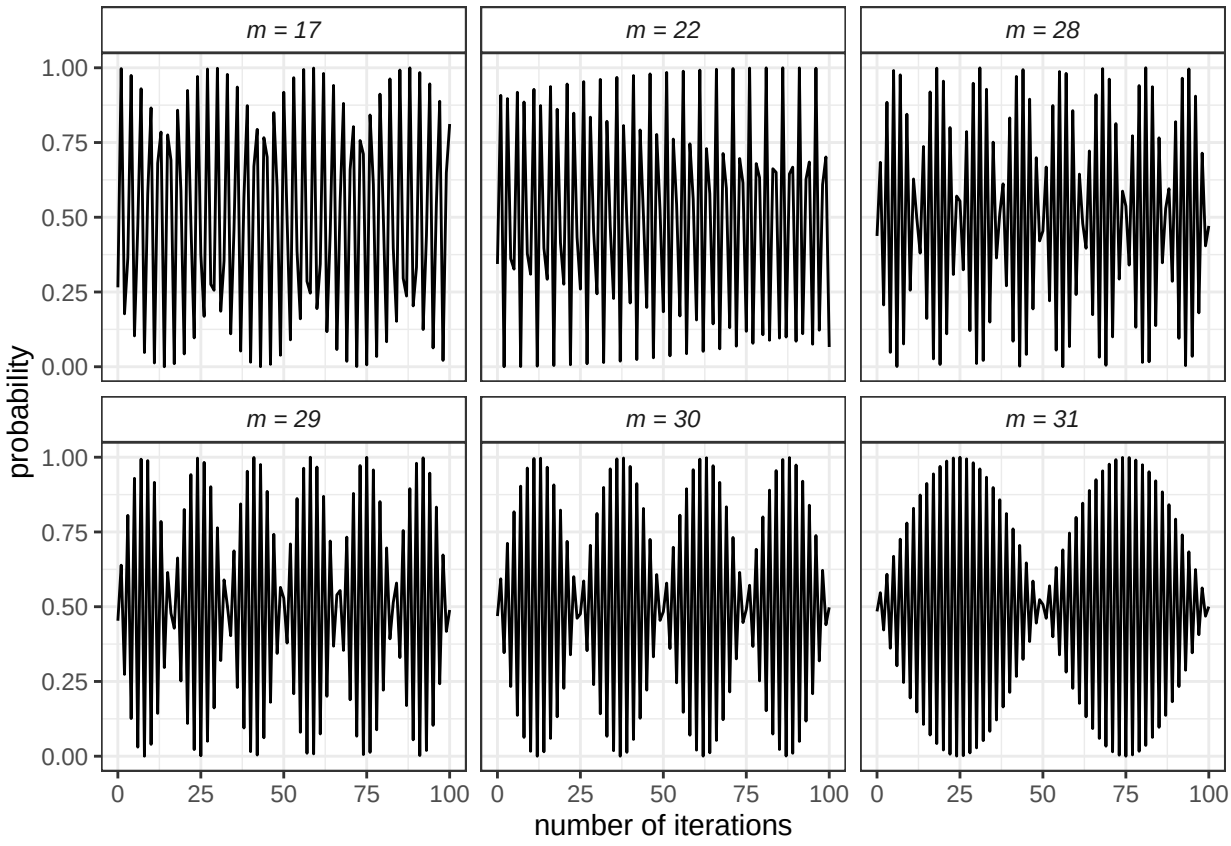


Figure 3. Probability of measuring a target basis state after each Grover iteration for a system that has 6 qubits and 17, 22, 28, 29, 30, and 31 target basis states.



probability amplitude of the target basis states (α_t), and the current iteration (i). Next, in each iteration i, we calculate the average probability amplitude ($\bar{\alpha}$) as the weighted sum of the probability amplitudes of the target and the other basis states. We use the average probability amplitude to update the probability amplitude of the target states (i.e., we reflect α_t about $\bar{\alpha}$), and calculate $P(n, m, i)$; the probability to successfully identify a target basis state after i Grover iterations in a system that has n qubits and m target basis states. If $P(n, m, i) \geq \rho$, we return i as the number of required iterations. Otherwise, if $P(n, m, i) < \rho$, we update the probability amplitude of the regular states and increment i. We continue until the required number of iterations has been identified.

Algorithm 9: Procedure ($\texttt{GI}\rho$) to determine the number of iterations required by Grover's algorithm to measure a target basis state with probability of at least ρ in a system that has n qubits and m target basis states.

initialize $\alpha_o = (\sqrt{2^n})^{-1}$, $\alpha_t = -\alpha_o$, and $i = 1$;
do
  $\bar{\alpha} = 2^{-n}((2^n - m)\alpha_o + m\alpha_t)$;
  $\alpha_t = \alpha_t - 2\bar{\alpha}$;
  $P(n, m, i) = \alpha_t^2$;
  if $P(n, m, i) \geq \rho$ **then**
    return i;
  $\alpha_o = 2\bar{\alpha} - \alpha_o$;
  $i = i + 1$;
while true;

Using procedure $\texttt{GI}\rho$, we perform an experiment to determine the number of Grover iterations required to measure a target basis state with 95 percent confidence for various values of n and m. Figure 4 presents the results of this experiment and compares the performance of $\texttt{GI}\rho$ (with $\rho = 0.95$) and approximation $I(n, m)$. For each value of $n : n \in \{6, 10, 16\}$, Figure 4 shows the number of Grover iterations for both approaches (top panel) as well as the probability to measure a target basis state (bottom panel). From Figure 4, it becomes clear that approximation $I(n, m)$ cannot be used with confidence for all values of m. In addition, Figure 4 also shows that, even if we use $\texttt{GI}\rho$ to determine the required number of Grover iterations, it may be impossible to achieve an acceptable level of confidence to measure a target basis state. For instance, if half of the basis states are target basis states, the probability to measure a target basis state always equals 0.5, regardless of the number of Grover iterations that are used (i.e., in this case, using Grover's algorithm is equivalent to randomly guessing a target basis state). This also illustrates that Grover's algorithm itself does not belong to quantum complexity class BQP (Bounded-error Quantum Polynomial time; the quantum counterpart of complexity class BBP), as it cannot guarantee that a target basis state is measured (i.e., that a solution is found) with an error probability of at most $1/3$.

We can partly remedy the problems of approximation $I(n, m)$ by using, for instance, approximation $\max(1, \lfloor \pi 4^{-1} \sqrt{m^{-1} 2^n} \rfloor)$. This is shown in Figure 5, that compares the performance of approximation $I(n, m) = \lceil \pi 4^{-1} \sqrt{m^{-1} 2^n} \rfloor$ with that of $\max(1, \lfloor \pi 4^{-1} \sqrt{m^{-1} 2^n} \rfloor)$. Investigating better approximations(that minimize the number of Grover iterations while maximizing the probability to measure a target basis state) is left as a di-

Figure 4. Comparison of the performance of approximation $\pi 4^{-1}\sqrt{m^{-1}2^n}$ (black line; implemented as $I(n,m)$) and procedure $\texttt{GI}\rho$ (gray line; with $\rho = 0.95$). For each value $n : n \in \{6, 10, 16\}$, the top panel shows the number of Grover iterations whereas the bottom panel shows the probability to measure a target basis state.

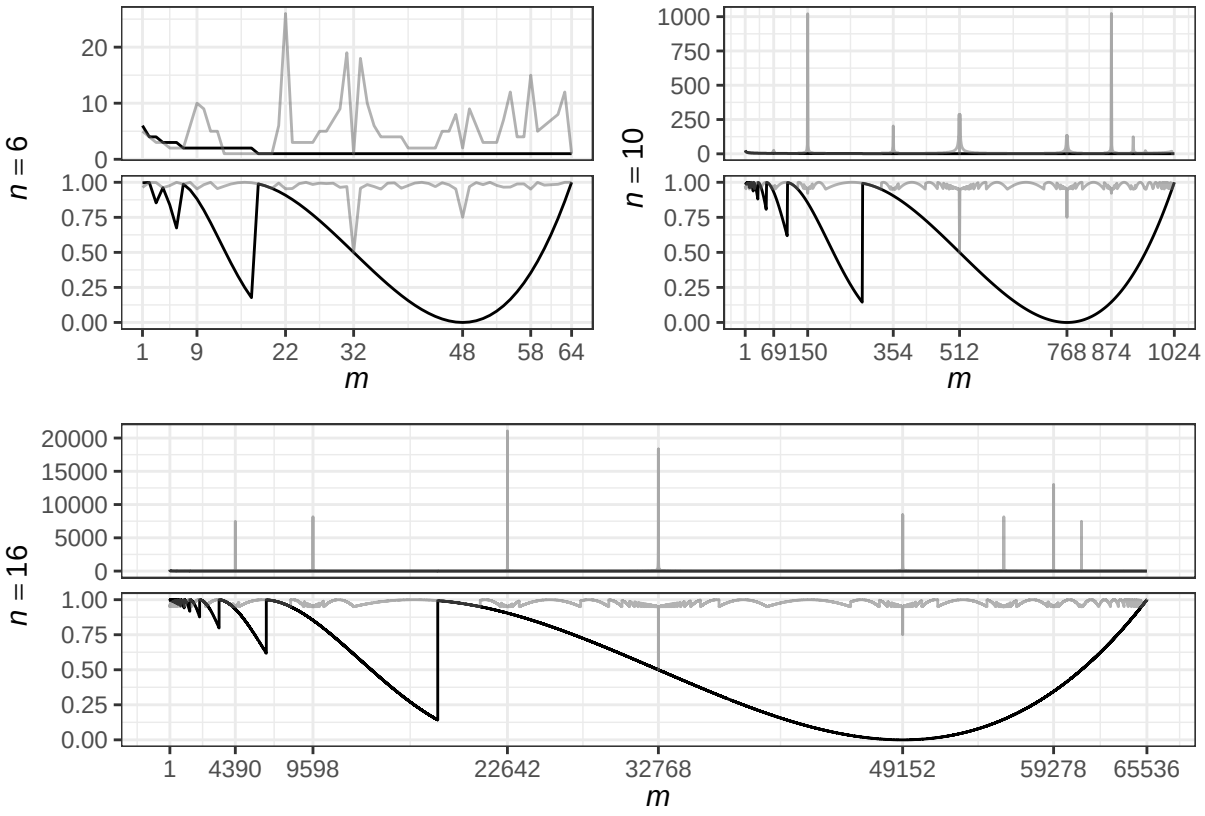
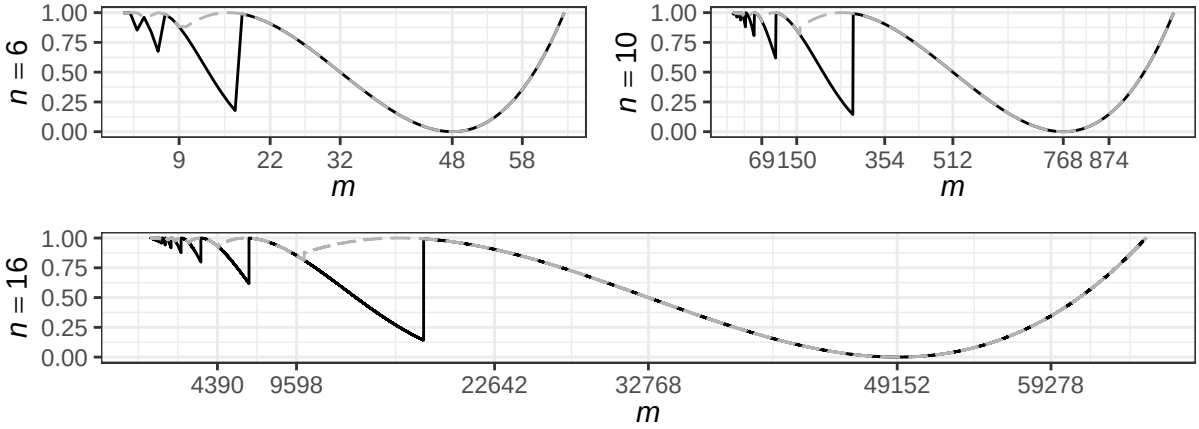


Figure 5. Comparison of the performance of approximation $\lceil \pi 4^{-1}\sqrt{m^{-1}2^n} \rceil$ (black full line) and approximation $\max(1, \lfloor \pi 4^{-1}\sqrt{m^{-1}2^n} \rfloor)$ (gray dashed line). For each value of $n : n \in \{6, 10, 16\}$, the figure shows the probability to measure a target basis state for various values of m.



rection for future research.

C. Performance of GUM

The goal of GUM is to identify one of the m valid solutions in a system that has 2^n solutions. An outline of the procedure is provided by Algorithm 1 in the main paper. The expected number of calls to function $f_{\mathbf{x}}$ required by GUM is given by:

$$\sum_{i=0}^{n} \left((i+1) \sum_{j=0}^{i} 2I(n, 2^{(n-j)}) \right) (1 - \phi_{i-1}) P(n, m, I(n, 2^{(n-i)})), \tag{1}$$

where ϕ_i is the probability of having measured a valid solution after i GUM iterations and can be obtained through the following recursive relationship (with $\phi_{-1} = 0$):

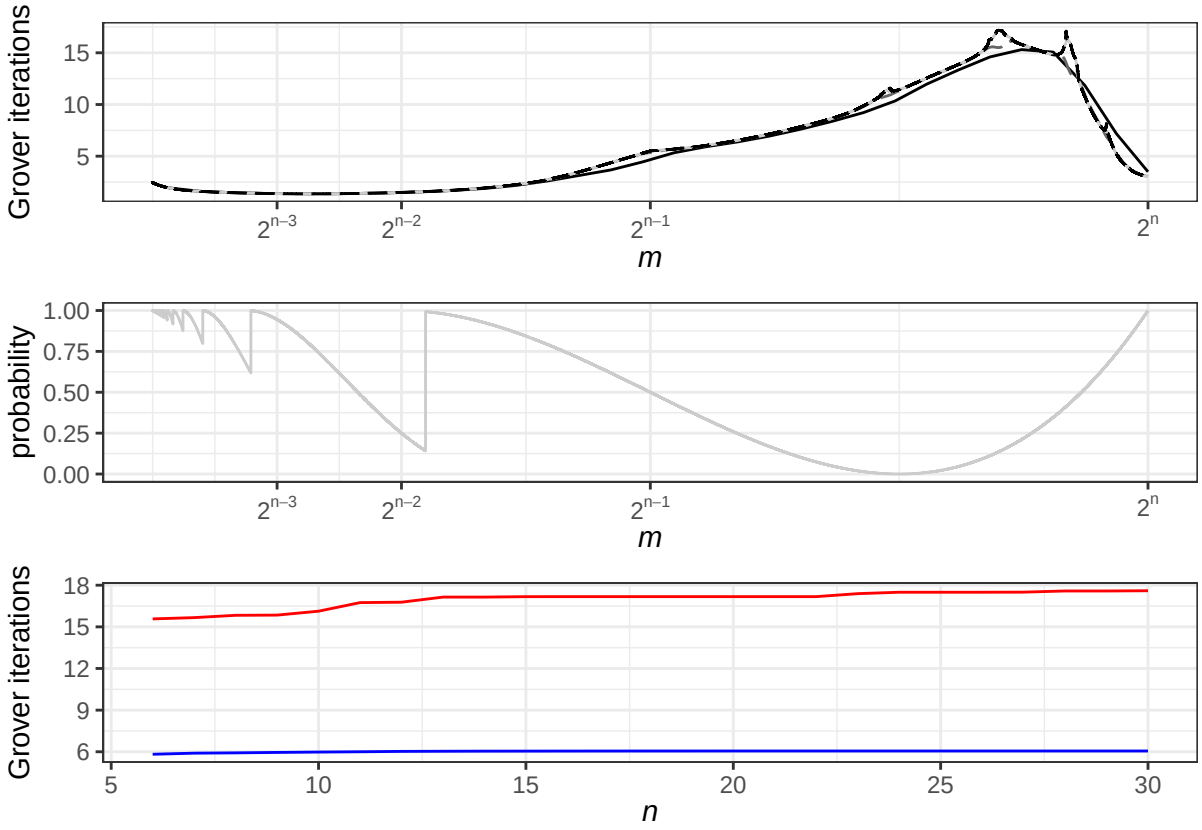
$$\phi_i = \phi_{i-1} + (1 - \phi_{i-1}) P(n, m, I(n, 2^{(n-i)})). \tag{2}$$

If we are lucky, GUM identifies a valid solution in the first run of Grover's algorithm (i.e., when assuming $\widehat{m} = 2^n$). In this case, we require two times $I(n, 2^n)$ calls to $f_{\mathbf{x}}$ (once to flip the phase of all valid solutions and once to uncompute) and one call to verify whether the measured solution is valid. If, on the other hand, no valid solution was found in the first run (with probability $1 - P(n, m, I(n, 2^n))$), we perform a second run of Grover's algorithm in which $\widehat{m} = 2^{(n-1)}$. If we find a solution in this run (with probability $P(n, m, I(n, 2^{(n-1)}))$), we performed $1 + 2I(n, 2^n)$ calls to $f_{\mathbf{x}}$ in a previous run, and will perform another $1 + 2I(n, 2^{(n-1)})$ calls in this run. In total, we have performed $2 + \sum_{j=0}^{1} 2I(n, 2^{(n-j)})$ calls to function $f_{\mathbf{x}}$. In general, after i runs, we have performed $(i+1) \sum_{j=0}^{i} 2I(n, 2^{(n-j)})$ calls to function $f_{\mathbf{x}}$. In the worst case, no valid solution can be found, and GUM requires $(n+1) + \sum_{i=0}^{n} 2I(n, 2^{(n-i)}) \approx O(\sqrt{2^n})$ function calls.

To further investigate the performance of GUM, we perform a computational experiment. The results of the experiment are summarized in Figure 6. The top panel of Figure 6 shows the number of operations (divided by $\sqrt{m^{-1}2^n}$) required by GUM for various values of m and n (note that we divide by $\sqrt{m^{-1}2^n}$ to show that GUM requires $O(\eta\sqrt{m^{-1}2^n})$ operations, where η is a constant). We observe that the expected number of operations is rather constant if $m < 2^{(n-2)}$. Only for values of $m > 2^{(n-2)}$, the expected number of operations starts to increase. This is not surprising as approximation $I(n, m)$ (that is used by GUM to determine the number of Grover iterations) is inaccurate if $m > 2^{(n-2)}$ (see also the middle panel of Figure 6 that shows the probability that Grover's algorithm is able to measure any of the m solutions if $I(n, m)$ Grover iterations are used). To resolve this problem, we can use $f_{\mathbf{x}}$ to evaluate, for example, 20 random solutions before running GUM. If we have a probability of at least $2^{(n-2)} = 0.25$ to find a valid solution, with 20 trials we are at least 99.68 percent certain to find a valid solution. If no valid solution was found after 20 trials, we can safely assume there are (far) less than $2^{(n-2)}$ valid solutions, and we run GUM. Last but not least, the bottom panel of Figure 6 shows the average and the maximum number of expected operations (divided by $\sqrt{m^{-1}2^n}$) required by GUM over all values of m, for various values of n. We observe that the average and the maximum number of operations converge as n increases. In addition, the expected number of operations (divided by $\sqrt{m^{-1}2^n}$) shown in the top panel also converges as n increases. As a result, we conjecture that GUM requires $O(\sqrt{m^{-1}2^n})$ operations to find any of the m valid solutions in a set of 2^n solutions (rather than $O(\sqrt{2^n})$ operations).

As we have seen in Section B, approximation $I(n, m)$ cannot always be used with confidence to identify a valid solution for all values of m. Procedure GUM, however, relies on approximation $I(n, m)$ to determine the number of Grover iterations to be used in each run of Grover's algorithm. Therefore, we might wonder whether GUM is able to always identify a valid solution if such a solution exists? Before answering this question, note that it is certainly possible to construct an example where GUM only has a small probability to find a valid solution. For non-trivial values of n, however, one can be quite confident that GUM will return a valid solution. To explain why, we observe two facts. First, if there are m valid solutions, the probability of measuring a valid solution after i runs can be obtained using Equation 2. Second, the probability amplitude of the target basis states can be approximated by a sine function (see also Figure 6 in the main paper). As a result, if we assume that $\widehat{m} = 2^n$, $\widehat{m} = 2^{(n-1)}$, ..., $\widehat{m} = 2$, $\widehat{m} = 1$ is a random sample of that sine function, $P(n, m, I(n, \widehat{m}))$ equals 0.5 on average, and hence, ϕ_n approaches 1, even for small n. This is also illustrated in Table 1. Table 1 presents ϕ_i after each of the $(n+1)$ runs of Grover's algorithm for a system that has $n = 6$ qubits, and for various values of m (the values of m that are used in the table were selected because they have a low probability to yield a valid solution, or because they require a large number of Grover iterations in order to ensure a probability of at least 0.95 to measure a valid solution; see also Figure 4). Table 1 also reports the expected total number of Grover iterations required to measure a valid solution (last column). It is clear that, regardless the value of m, we are almost certain to measure a valid solution. In the worst case, $m = 59$, and we have to perform $(n + 1) = 7$ runs of Grover's algorithm to be 97.9 percent certain to measure a valid solution (this is the smallest value of ϕ_n over all values of m). In other words, even for small n, we are almost certain to measure a valid solution. If, on the other hand, no valid solution exists (i.e., if $m = 0$), we are 100 percent certain that no valid solution will be found after performing 7 runs (requiring a total of $1 + 1 + 2 + 2 + 3 + 4 + 6 = 19$ Grover iterations). Note that, if $m > 0$, we expect to perform far less

Figure 6. The top panel shows the expected number of operations divided by $\sqrt{m^{-1}2^n}$ required by GUM for various values of m, and for $n = 6$ (black full line), $n = 10$ (dark gray dashed line), $n = 16$ (black dashed line), and $n = 22$ (gray full line). The middle panel shows the probability to measure one of the m valid solutions using Grover's algorithm if $I(n, m)$ Grover iterations are used. The bottom panel shows the average (blue) and maximum (red) number of expected operations divided by $\sqrt{m^{-1}2^n}$ required by GUM over all values of m, for various values of n.



Grover iterations. This illustrates why GUM may require far less than $O(\sqrt{m^{-1}2^n})$ calls to function $f_{\mathbf{x}}$ (i.e., its expected performance is far better than its worst-case performance).

Table 1. Probability to measure one of the m valid solutions after i runs of Grover's algorithm in a system that has 6 qubits and expected number of Grover iterations (E [GI]) required to successfully measure a valid solution, for various values of m. The worst performance (over all values of m) corresponds to the case where $m = 59$, with a probability of 0.979 to measure a valid solution after $(n + 1)$ runs; after a total of 19 Grover iterations (indicated in red).

i	0	1	2	3	4	5	6	
$\widehat{m}$	64	32	16	8	4	2	1	
$I(n, \widehat{m})$	1	1	2	2	3	4	6	
m	ϕ_i							E [GI]
0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	19.00
1	0.135	0.251	0.509	0.678	0.868	0.976	1.000	5.981
3	0.371	0.604	0.916	0.982	1.000	1.000	1.000	2.644
6	0.646	0.875	1.000	1.000	1.000	1.000	1.000	1.605
9	0.836	0.973	0.997	1.000	1.000	1.000	1.000	1.229
17	0.997	1.000	1.000	1.000	1.000	1.000	1.000	1.003
22	0.908	0.991	0.991	0.991	0.999	0.999	1.000	1.158
31	0.547	0.795	0.881	0.931	0.973	0.983	0.988	2.290
32	0.500	0.750	0.875	0.938	0.969	0.984	0.992	2.508
33	0.453	0.701	0.874	0.947	0.968	0.988	0.996	2.690
48	0.000	0.000	0.750	0.938	0.984	0.984	0.996	4.770
58	0.354	0.583	0.583	0.583	0.719	0.969	0.981	5.511
59	0.436	0.682	0.689	0.696	0.745	0.921	0.979	4.829
64	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

D. Solving the binary knapsack problem using Grover's algorithm

In order to use Grover's algorithm to effectively solve the binary knapsack problem, we cannot rely on an oracle function (that has complete knowledge of all valid solutions). Instead, we have to implement function $f_{\mathbf{x}}$ on a quantum computer. The implementation of $f_{\mathbf{x}}$ as a quantum circuit is presented in Figure 7. The circuit may be divided into three parts (indicated by the red dashed lines in the circuit). In the first two parts, we determine whether knapsack $\mathbf{x}$ satisfies the weight- and value constraints (constraints 4 and 6 of the binary knapsack problem formulation available in the main paper). In a third part, we obtain the output of function $f_{\mathbf{x}}$. In what follows, we use an example to demonstrate the dynamics of the circuit. In the example, we have a pool of three items and a knapsack that has maximum weight capacity $W = 4$. The weights and values of the items are listed in Table 2. Since we are working with qubits, all weights and values need to be converted to binary values. Let b denote the number of qubits that are required to represent the largest weight or value used in the example. In our example, $b = 3$, and the qubits that represent the maximum weight capacity are initialized as follows: $\left|c_1^W\right\rangle = |1\rangle$, $\left|c_2^W\right\rangle = |0\rangle$, and $\left|c_3^W\right\rangle = |0\rangle$ (with $\mathbf{c}^W = \left\{c_1^W, \ldots, c_b^W\right\}$). The binary conversions of the weights and values of the items are listed in Table 2, where $\mathbf{c}^w = \left\{\mathbf{c}_1^w, \ldots, \mathbf{c}_n^w\right\}$, $\mathbf{c}^v = \left\{\mathbf{c}_1^v, \ldots, \mathbf{c}_n^v\right\}$, $\mathbf{c}_i^w = \left\{c_{i1}^w, \ldots, c_{ib}^w\right\}$, and $\mathbf{c}_i^v = \left\{c_{i1}^v, \ldots, c_{ib}^v\right\}$. In our example, $\left|\mathbf{c}^W\right\rangle$, $|\mathbf{c}^w\rangle$, and $|\mathbf{c}^v\rangle$ are initialized as $|100\rangle$, $|010011010\rangle$, and $|011001010\rangle$, respectively. Next, imagine that we use the circuit of function $f_{\mathbf{x}}$ to evaluate knapsack $\mathbf{x} = \{1, 0, 1\}$. In this case, $|\mathbf{x}\rangle$ is initialized as $|101\rangle$. In a first part of the

Figure 7. Circuit of function $f_{\mathbf{x}}$ that can be used to verify whether a knapsack solution $\mathbf{x}$ has a weight smaller than W and a value of at least V.

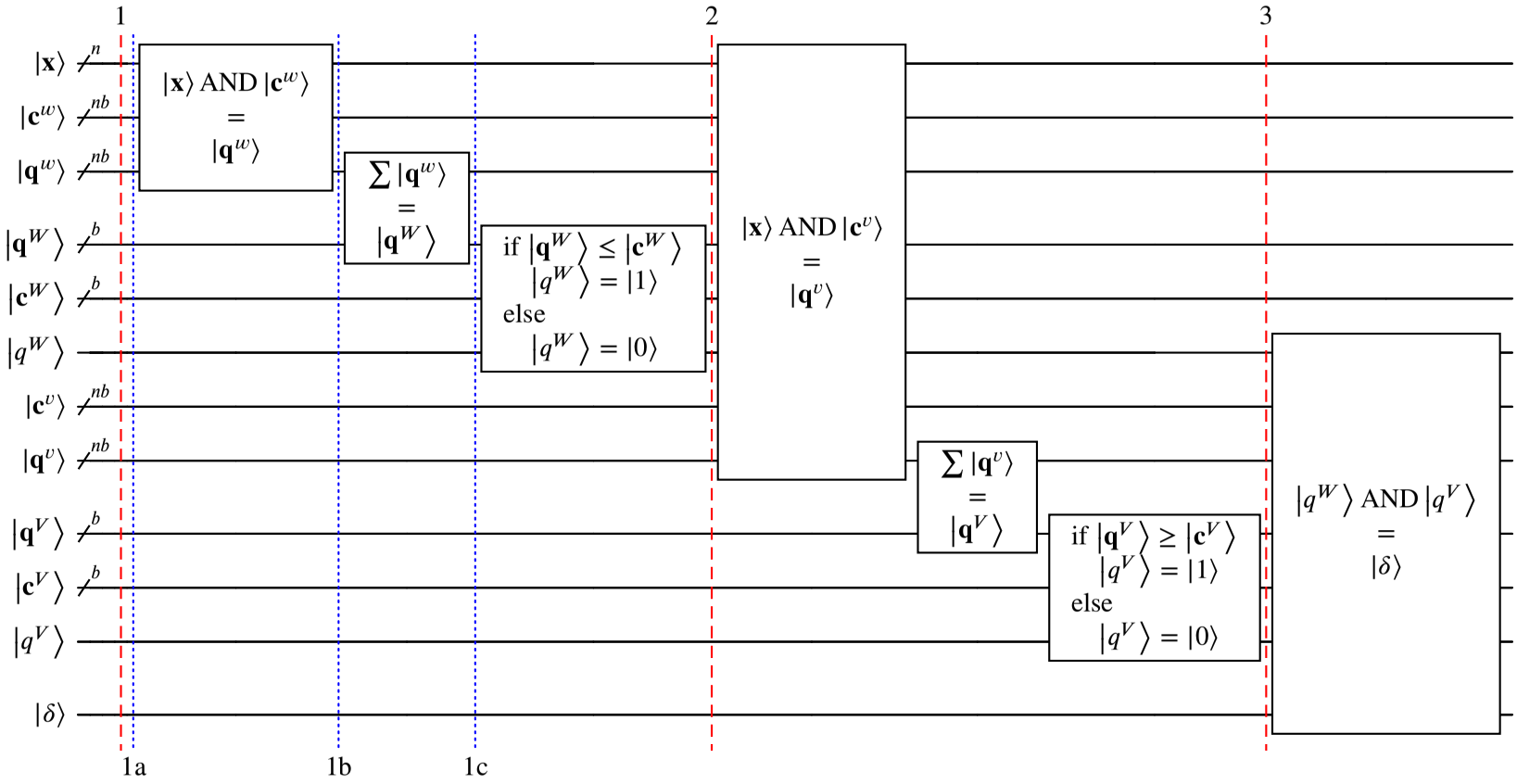


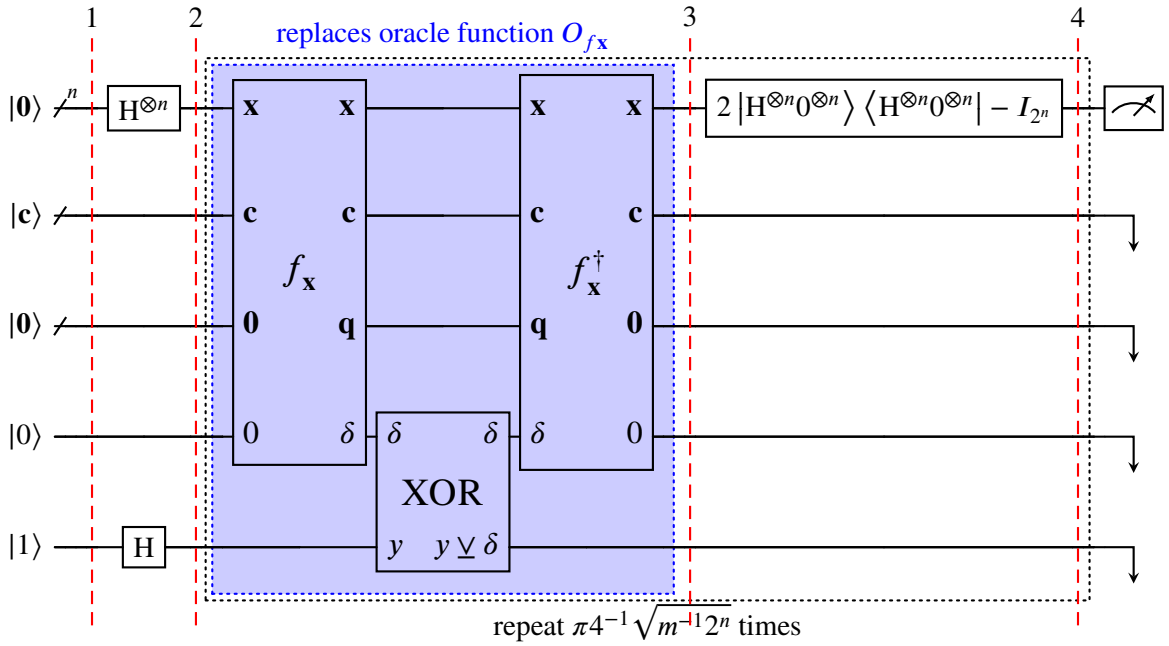
Table 2. Data for example knapsack problem with $n = 3$ items.

i	w_i	c^w_{i1}	c^w_{i2}	c^w_{i3}	v_i	c^v_{i1}	c^v_{i2}	c^v_{i3}
1	2	0	1	0	3	0	1	1
2	3	0	1	1	1	0	0	1
3	2	0	1	0	2	0	1	0

circuit, we distinguish three steps (indicated by the blue dotted lines in the circuit). In a first step (1a), we determine how much weight each item contributes to knapsack $\mathbf{x}$. The contributed weight of the items is represented by $\mathbf{q}^w = \{\mathbf{q}^w_1, \ldots, \mathbf{q}^w_n\}$, with $\mathbf{q}^w_i = \{q^w_{i1}, \ldots, q^w_{ib}\}$. $|\mathbf{q}^w\rangle$ is initialized as $|\mathbf{0}\rangle$ and is used to store the result of an AND operation on $|\mathbf{x}\rangle$ and $|\mathbf{c}^w\rangle$. In our example, one can verify that $|\mathbf{q}^w\rangle = |010000010\rangle$. In a second step (1b), we obtain the weight of knapsack $\mathbf{x}$ itself (represented by $\mathbf{q}^W = \{q^W_1, \ldots, q^W_b\}$). $|\mathbf{q}^W\rangle$ is initialized as $|\mathbf{0}\rangle$ and is used to store the result of $\sum_{i=1}^{n} |\mathbf{q}^w_i\rangle$. To perform the aforementioned addition, we can create our own adder circuits or use a more efficient circuit that was proposed by Draper (2002). In our example, $|\mathbf{q}^W\rangle = |100\rangle$. As a last step (1c), we verify whether knapsack $\mathbf{x}$ satisfies the weight constraint. This can be done using the quantum equivalent of a digital magnitude comparator (see, for example, Xia et al (2018)). The digital magnitude comparator compares the weight of knapsack $\mathbf{x}$ (represented by $\mathbf{q}^W$) and the maximum knapsack weight (represented by $\mathbf{c}^W$), and outputs 1 if a smaller-than-or-equal-to condition is met. The result is stored in a qubit that is initialized as $|q^W\rangle = |0\rangle$. In our example, the weight constraint is satisfied, and $|q^W\rangle$ evaluates to $|1\rangle$. This concludes the first part of the circuit. The second part of the circuit is almost identical to the first part, and verifies whether the value of the knapsack (stored in $|\mathbf{q}^V\rangle = \sum_{i=1}^{n} |\mathbf{q}^v_i\rangle$, where $\mathbf{q}^V = \{q^V_1, \ldots, q^V_b\}$, $\mathbf{q}^v = \{\mathbf{q}^v_1, \ldots, \mathbf{q}^v_n\}$, and $\mathbf{q}^v_i = \{q^v_{i1}, \ldots, q^v_{ib}\}$) is at least equal to value V (stored in $|\mathbf{c}^V\rangle$, where $\mathbf{c}^V = \{c^V_1, \ldots, c^V_b\}$). If this is the case, the value constraint is satisfied, and $|q^V\rangle = |1\rangle$. If, however, knapsack $\mathbf{x}$ does not satisfy the value constraint, then $|q^V\rangle = |0\rangle$. In a last part of the circuit, we obtain the outcome of function $f_{\mathbf{x}}$ itself (stored in a qubit that is initialized as $|\delta\rangle = |0\rangle$). If both constraints are satisfied, $|\delta\rangle = |1\rangle$ (and $|\delta\rangle = |0\rangle$ otherwise).

Figure 8 presents the circuit of Grover's algorithm equipped with function $f_{\mathbf{x}}$ (the steps of the algorithm are indicated by red dashed lines). In this circuit, $\mathbf{x}$, $\mathbf{c}$, $\mathbf{q}$, and δ are the qubits that are used to store the decision variables, the constant values that are used by function $f_{\mathbf{x}}$, the outcomes of operations performed by function $f_{\mathbf{x}}$, and the result of function $f_{\mathbf{x}}$, respectively. In addition, y is a qubit that is used to flip the phase of valid solutions. $|y\rangle$ is initialized as $|1\rangle$. The initialization of $|\mathbf{c}\rangle$ depends on the problem instance. All other qubits are initialized as $|0\rangle$. The gates grouped in blue replace oracle gate $O_{f_{\mathbf{x}}}$ in the original circuit of Grover's algorithm (see Figure 3 in the main paper). These gates perform three operations. First, function $f_{\mathbf{x}}$ evaluates whether $\mathbf{x}$ is valid, and stores the result in $|\delta\rangle$. Next, a XOR operation is used to map $|\delta y\rangle$ to $|\delta(y \veebar \delta)\rangle$. After that, function $f^{\dagger}_{\mathbf{x}}$ applies all gates of function $f_{\mathbf{x}}$ in reverse order. By doing so, function $f^{\dagger}_{\mathbf{x}}$ reinitializes $|\mathbf{q}\rangle$ and $|\delta\rangle$ (as $|\mathbf{0}\rangle$ and $|0\rangle$, respectively) such that they are ready for a next Grover iteration (note that $\mathbf{c}$ and y are unaffected and remain in state $|\mathbf{c}\rangle$ and $|-\rangle$ until they are discarded at the end of the procedure). This process ensures that the circuit is reversible and is also known as "uncomputation" (refer to Nielsen and Chuang (2010) for more details).

Figure 8. Circuit of Grover's algorithm equipped with function $f_{\mathbf{x}}$.



E. Directions for future research

Below, we provide a list of future research topics:

1. Improve the $I(n,m) = \left\lceil \pi 4^{-1}\sqrt{m^{-1}2^n} \right\rceil$ approximation; identify approximations that minimize the number of Grover iterations while maximizing the probability to measure a valid solution (see also Section B). Note that the resulting approximation may be a hybrid procedure that first verifies a number of solutions using a classical method before trying to use Grover's algorithm to measure a valid solution (to counter the effect that Grover's algorithm may require many Grover iterations even if m is large; see also Section B).

2. Verify the potential of running Grover's algorithm in parallel; investigate the trade-off between running multiple instances of Grover's algorithm and reducing the number of Grover iterations that are used by each instance of Grover's algorithm (i.e., whereas running multiple instances of Grover's algorithm in parallel increases the probability to measure a valid solution, reducing the number of Grover iterations for each instance of the algorithm decreases the probability; see also Section 10 in the main paper).

3. Verify the potential of running `GUM` (and by extension `HBB`, `RAP`, and `RAHBB`) in parallel (see also Section 10 in the main paper).

4. Investigate closed-form (sine) approximations for $P(n,m,I)$ that may be used instead of Algorithm 7 (see also Section B of the supplementary material and Appendix A in the main paper).

5. Investigate approximation methods to assess the performance of quantum algorithms as well as approximation methods for determining m_V. See also Appendix B in the main paper.

6. Investigate the potential of procedures such as HBB and RAP to solve discrete optimization problems other than the binary knapsack problem; identify problems for which procedures such as HBB and RAP outperform the best classical algorithms.

7. Develop dedicated quantum algorithms (rather than general-purpose procedures such as HBB and RAP) that can be used for solving specific optimization problems.

8. Solve continuous optimization problems rather than only discrete optimization problems (i.e., generalize procedures such as HBB and RAP).

9. Investigate the use of other Grover-based algorithms and/or extensions of Grover's algorithm such as quantum counting, nested quantum search, amplitude amplification, etc.

10. Investigate quantum algorithms that do not rely on Grover's algorithm such as quantum walk algorithms or algorithms that use adiabatic quantum computation.

References

Aaronson, S. 2018. *Introduction to Quantum Information Science*. Lecture Notes, UT Austin.

Aharonov, D. 2003. A simple proof that Toffoli and Hadamard are quantum universal, unpublished preprint at https://doi.org/10.48550/arXiv.quant-ph/0301040.

Deutsch, D. 1989. Quantum computational networks. *Proc. R. Soc. Lond. A*, **425**(1868), 73–90.

Dirac, P. A. M. 1939. A new notation for quantum mechanics. *Math. Proc. Camb. Philos. Soc.*, **35**(3), 416–418.

Draper, T. G. 2000. Addition on a quantum computer, unpublished preprint at https://doi.org/10.48550/arXiv.quant-ph/0008033.

Nielsen, M., Chuang, I. 2010. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, Cambridge.

Xia, H., Li, H., Zhang, H., Liang, Y., and Xin, J. 2018. An efficient design of reversible multi-bit quantum comparator via only a single ancillary bit. *Int. J. Theor. Phys.*, **57**(12), 3727–3744.