Constrained mixers for QAOA

Franz G. Fuchs†, Kjetil Olsen Lye†, Halvor Møll Nilsen†, Alexander J. Stasik†, and Giorgio Sartor†

†SINTEF AS, Department of Mathematics and Cybernetics, Oslo, Norway

March 14, 2022

Abstract

The quantum approximate optimization algorithm/quantum alternating operator ansatz (QAOA) is a heuristic to find approximate solutions of combinatorial optimization problems. Most literature is limited to quadratic problems without constraints. However, many practically relevant optimization problems do have (hard) constraints that need to be fulfilled. In this article we present a framework for constructing mixing operators that restrict the evolution to a subspace of the full Hilbert space given by these constraints. We also describe algorithms for efficient decomposition into basis gates. The underlying mathematical structure reveals more of how mixers work and how one can minimize their cost in terms of number of CNOT gates, particularly when Trotterization taken into account.

1 Introduction

The quantum approximate optimization algorithm (QAOA) [2], and its generalization, the quantum alternating operator ansatz (also abbreviated as QAOA) [6] is a meta-heuristic for solving combinatorial optimization problems that can utilize gate based quantum computers and possibly outperform purely classical heuristics. Typical examples that can be tackled are quadratic (binary) optimization problems of the form

\[ x^* = \arg \min_{x \in \{0,1\}^n} f(x), \quad f(x) = x^T Q_f x + c_f, \quad \text{such that} \quad g(x) = x^T Q_g x + c_g = 0 \]  

where \( Q_f, Q_g \in \mathbb{R}^{n \times n} \) are symmetric \( n \times n \) matrices. For binary variables \( x \in \{0,1\} \), any linear part can be absorbed into the diagonal of \( Q_f \) and \( Q_g \). In this article we focus on the case where the constraint is given by a feasible subspace as defined in the following.

Definition 1 (Constraints given by indexed computational basis states). Let \( \mathcal{H} = (\mathbb{C}^2)^\otimes n \) the Hilbert space for \( n \) qubits, which is spanned by all computational basis states \( |z_j \rangle \), i.e., \( \mathcal{H} = \text{span}\{ |z_j \rangle, 1 \leq j \leq 2^n, z_j \in \{0,1\}^n \} \). Let

\[ B = \{ |z_j \rangle, \quad j \in J, \quad z_j \in \{0,1\}^n \}, \]

the subset of all computational basis states defined by an index set \( J \). This corresponds to

\[ g(x) = \sum_{j \in J} \sum_i (x_i - (z_j)_i)^2, \]

which is a quadratic constraint.

There is a well-established connection of quadratic (binary) optimization problems to Ising models, see e.g. [11], that allows to directly translate these problems to the QAOA. The general form of QAOA is given by

\[ |\gamma, \beta \rangle = U_M(\beta_p)U_P(\gamma_p) \cdots U_M(\beta_1)U_P(\gamma_1) |\phi_0 \rangle, \]

where a phase separating and mixing operator is alternately applied \( p \) times. As defined in [6] the requirements for these operators are as follows
• a phase separating operator $U_P(\gamma)$ depending on the objective function $f$,

• a mixing operator $U_M(\beta)$ with the following properties,
  
  - $U_M$ does not commute with $U_P$, i.e., $[U_M(\beta), U_P(\gamma)] \neq 0$, for almost all $\gamma, \beta \in \mathbb{R}$,
  
  - $U_M$ preserves the feasible subspace as given in Definition 1, i.e., $\text{Sp}(B)$ is an invariant subspace of $U_M$,
    
    $$U_M(\beta) |v\rangle \in \text{Sp}(B), \quad \forall |v\rangle \in \text{Sp}(B), \forall \beta \in \mathbb{R},$$

  - $U_M$ provides transitions between all pairs of feasible states, i.e., for each pair $x, y \exists \beta^* \in \mathbb{R}$ and $\exists r \in \mathbb{N} \cap \{0\}$, such that
    
    $$\left| x U_M(\beta^*) \cdots U_M(\beta^*) \right| y \rangle > 0, \forall \text{comp. basis states } |x\rangle, |y\rangle \in B.$$

• a feasible initial state $|\phi_0\rangle$ is an eigenvector of $U_M$.

**Remark 1 (Repeated mixers).** If $U_M$ is the exponential of a Hermitian matrix, the parameter $r$ does not matter as it can be absorbed as a re-scaling of $\beta$. However, if $U_M$ is Trotterized, i.e., the product of exponentials of non-commuting Hermitian matrices, this can lead to missing transitions. In this case $r > 1$ can again provide these transitions. It is therefore suggested as useful in [6] to repeat mixers within one mixing step. Out of this reason, we will treat the cost of Trotterized mixers including the necessary repetitions to provide transitions for all feasible states.

## 2 Related work

There are two principal ways to take constraints into account with the QAOA. The standard approach to enforce the constraints in Equation 1 is to add them to the target function with the help of a penalty (Lagrange multiplier) $\lambda$, leading to

$$x^* = \arg \min_{x \in \{0, 1\}^n} \left( f(x) + \lambda g(x) \right).$$

This approach is popular, since it is straightforward to define a phase separating Hamiltonian for $f(x) + \lambda g(x)$. Example applications are the tail-assignment problem [12], the Max-k-cut problem [3], graph coloring problems, and the traveling sales person problem [7]. A downside of this approach is that also infeasible solutions are possible outcomes for approximate solvers like QAOA. This has also the drawback that the search is less efficient as if the evolution were confined to feasible configurations only. In addition, the quality of the obtained results turns out to be very sensitive to the chosen values of the hyper parameter $\lambda$ in practice. On one hand, $\lambda$ should be chosen large enough such that the lowest eigenstates of $H_P$ correspond to feasible solution. On the other hand, too large values of $\lambda$ mean that the resulting optimization landscape in the $\gamma$ has very high frequencies, which makes the problem hard to solve in practice. It is very challenging to find (the problem dependent) values for $\lambda$ that balance the influence of the cost/objective function and the penalty terms [13].

For QAOA, a second approach is to define mixers that have zero probability to go from a feasible state to an infeasible one. This avoids the hyper parameter that is necessary in the above approach. However, it is generally more challenging to devise mixers that take into account constraints. The most prominent example in literature is the XY-mixer [7, 6, 13] which constrains evolution to states with non-zero overlap with "one-hot" states. One-hot states are computational basis states with exactly one entry equal to one. For instance $|0001\rangle$ and $|101000\rangle$ are one-hot states, while $|00\rangle$ and $|110\rangle$ are not. The name XY mixer comes from the related XY-Hamiltonian [10]. The mixers derived in literature follow the intuition of physicists to use "hopping" terms.

QAOA can be viewed as a discretized version of quantum annealing. In quantum annealing enforcing constraints via penalty terms is particularly "harmful" since they often require all-to-all connectivity.
of the qubits [8]. The authors in [9] therefore introduce driver Hamiltonians that commute with the constraints of the problem. This bears similarities with and actually inspired the approaches in [7, 6].

The main contributions of this article are:

- A general framework to construct mixers restricted to a set of computational basis states, see Section 3.1.
- An analysis of the underlying mathematical structure, which is largely independent of the actual states, see Section 3.2.
- Efficient algorithms for decomposition into basis gates, see Section 3.3 and 3.5.
- Valid Trotterizations, which is not completely understood in literature, see Section 3.5.
- Improved efficiency of Trotterized mixers for "one-hot" states in Section 5.1.
- Discussion of the general case, exemplified in Section 5.2.

We start by describing the general framework.

3 Construction of constraint preserving mixers

In the following we will derive a general framework for mixers that are restricted to a subspace, given by certain basis states. For example, one may want to construct a mixer for five qubits that is restricted to the subspace \( \text{Sp}(|01001\rangle, |11001\rangle, |11110\rangle) \) of \( \mathbb{C}^2^n \), where \( \text{Sp}(B) \) denotes the linear span of \( B \).

In this section we will describe the conditions for preserving the feasible subspace, for providing transitions between all pairs of feasible states, as well as an efficient algorithm to decompose.

3.1 Conditions on the mixer Hamiltonian

**Theorem 1** (Mixer Hamiltonians for subspaces). Given a feasible subspace \( B \) as in Definition 1 and a real-valued transition matrix \( T \in \mathbb{R}^{|J| \times |J|} \). Then, for the mixer constructed via

\[
U_M(\beta) = e^{-i\beta H_M}, \quad \text{where} \quad H_M = \sum_{j,k \in J} (T)_{j,k} |x_j\rangle \langle x_k|,
\]

the following statements hold.

- If \( T \) is symmetric, the mixer is well defined and preserves the feasible subspace, i.e. condition (5) is fulfilled.
- If \( T \) is symmetric and for all \( 1 \leq j, k \leq |J| \) there exist an \( r \in \mathbb{N} \cup \{0\} \) (possibly depending on the pair) such that

\[
(T^r)_{j,k} \neq 0,
\]

then \( U_M \) provides transitions between all pairs of feasible states, i.e. condition (6) is fulfilled.

**Proof.** Well definedness.

Almost trivially \( H_M^\dagger \) is Hermitian if \( T \) is symmetric,

\[
H_M^\dagger = \sum_{j,k \in J} (T)_{j,k} |x_j\rangle \langle x_k| = \sum_{j,k \in J} (T)_{k,j} |x_k\rangle \langle x_j| = H_M.
\]

Since \( H_M \) is a Hermitian (and therefore normal) matrix there exists a diagonal matrix \( D \), with the entries of the diagonal as the (real valued) eigenvalues of \( H_M \), and a matrix \( U \), with columns given by the corresponding orthonormal eigenvectors. The mixer is therefore well defined through the convergent series

\[
e^{-itH_M} = \sum_{m=0}^{\infty} \frac{(-it)^m}{m!} H_M^m = U e^{-itD} U^\dagger.
\]
(a) The action of the Hamiltonian $H_{M,B}$ can be understood as the transition matrix $T$ acting upon the feasible basis states of $B = \{|x_j\rangle, \ j \in J\}$. Here, $E = \{|x_j\rangle\}_{j \in J}$.

(b) The Hamiltonian $H_{M,B} : \mathbb{C}^2 \rightarrow \mathbb{C}^2$ maps everything outside $\text{Sp}(B)$ to 0 and $\text{Sp}(B)$ onto itself.

Figure 1: Illustration of properties of Hamiltonians constructed with Theorem 1.

Reformulations.

We can rewrite $H_M$ in the following way

$$H_M : \ |y\rangle \mapsto \sum_{j,k \in J} (T)_{j,k} \langle x_j|y\rangle \ |x_k\rangle = ETE^T\ |y\rangle, \ |y\rangle \in \mathbb{C}^2,$$

where the columns of the matrix $E \in \mathbb{R}^{2^n \times |J|}$ consists of the feasible computational basis states, i.e., $E = \{|x_j\rangle\}_{j \in J}$, see Figure 1 for an illustration. Using that $TE = I \in \mathbb{R}^{|J| \times |J|}$ is the identity matrix, we have that

$$H_M^m = ET^mE^T = \sum_{j,k \in J} (T^m)_{j,k} \langle x_j| x_k\rangle, \ m \in \mathbb{N},$$

and Equation (11) can be written as

$$e^{-itH_M} = E \left( \sum_{m=0}^{\infty} \frac{(-it)^mT^m}{m!} \right) E^T.$$ 

Preservation of the feasible subspace.

Let $|v\rangle \in \text{Sp}(B)$. Using Equation (13) we know that

$$H_M^m |v\rangle = \sum_{j,k \in J} (T^m)_{j,k} \langle x_j| x_k\rangle \ |x_k\rangle = \sum_{j \in J} c_j \ |x_j\rangle \in \text{Sp}(B),$$

with coefficients $c_j \in \mathbb{C}$. Therefore, also $e^{-itH_M} |v\rangle \in \text{Sp}(B), t \in \mathbb{R}$, since it is a sum of these terms.

Transition between all pairs of feasible states.

For any pair of feasible computational basis states $|x_j\rangle$, $|x_{k^*}\rangle \in B$ we have that

$$f(t) = \langle x_{j^*}| U_M(t) |x_{k^*}\rangle = \langle x_{j^*}| \sum_{m=0}^{\infty} \frac{(-it)^m}{m!} \sum_{j,k \in J} (T^m)_{j,k} \langle x_j| x_k\rangle \ |x_{k^*}\rangle$$

$$= \sum_{m=0}^{\infty} \frac{(-it)^m}{m!} \langle x_j| x_{k^*}\rangle (T^m)_{j^*,k^*}.$$ 

It is enough to show that $f(t)$ is not the zero function. Since $f(t) : \mathbb{R} \rightarrow \mathbb{C}$ is an analytic function it has a unique extension to $\mathbb{C}$. Assume that $f$ is was indeed the zero function on $\mathbb{R}$, then the extension to $\mathbb{C}$ would also be the zero function and all coefficients of its Taylor series were zero. However, we assumed the existence of an $r \in \mathbb{N} \cup \{0\}$ such that $|\langle (T^r)_{j^*,k^*}\rangle| > 0$, and hence there exists a non-zero coefficient, which is a contradiction to $f$ being the zero function.
A natural question is how the statements in Theorem 1 depend on the particular ordering of the elements of $B$.

**Corollary 1** (Independence of the ordering of $B$). *Statements in Theorem 1 that hold for a particular ordering of computational basis states for a given $B$, hold also for any permutation $\pi : \{1, \cdots, |J|\} \to \{1, \cdots, |J|\}$, i.e., they are independent of the ordering of elements. For each ordering, the transition matrix $T$ changes according to $T_\pi = P_\pi^T P_\sigma$, where $P_\pi$ is the permutation matrix associated with $\pi$.

**Proof.** We start by pointing out that the inverse matrix of $P_\pi$ exists and can be written as $P_\pi^{-1} = P_{\pi^{-1}} = P_\sigma^T$. The resulting matrix $H_M$ is unchanged. Following the derivation in Equation (12), we have that $H_{M_\pi} = E_\pi T_\pi E_\pi^T$, where the columns of the matrix $E \in \mathbb{R}^{2^n \times |J|}$ consists of the permuted feasible computational basis states, i.e., $E_\pi = \{x_{\pi(j)}\}_{j \in J}$. Inserting $T = P_\pi^T P_\pi$ we have indeed $H_{M_\pi} = E_\pi T_\pi E_\pi^T = (E_\pi P_\pi^T T P_\pi E_\pi^T) = ETE^T = H_M$.

$T_\pi$ is symmetric if $T$ is. Assuming that $T^T = T$ we have that also

$$
(T_\pi)^T = (P_\pi^T T P_\pi)^T = P_\pi^T T^T P_\pi = P_\pi^T T P_\pi = T_\pi.
$$

If the condition in Equation (9) holds for $T$ than it also holds for $T_\pi$. Using $T^T = P_\pi^T T^T P_\pi$ we can show that Equation (9) holds for the permuted index pair $(\pi(j), \pi(k))$ for $T_\pi$ if it holds for $(j, k)$ for $T$. \hfill $\square$

In the following, if nothing else is remarked, computational basis states are ordered with respect to increasing integer value, e.g., $|001\rangle, |010\rangle, |111\rangle$.

Apart from special cases, there is a lot of freedom to choose the transition matrix $T$ that fulfills the conditions of Theorem 1. The entries of $T$ will heavily influence the circuit complexity, which will be investigated in Section 3.3. In addition, we have the following property which adds additional flexibility to develop efficient mixers.

**Corollary 2** (Properties of mixers). *For a given feasible subspace $Sp(B)$ let $U_{M,B}$ the mixer given by Theorem 1. For any subspace $Sp(C)$ with $Sp(B) \cap Sp(C) = \emptyset$ or equivalently $B \cap C = \emptyset$, also $U_M = U_{M,B} U_{M,C}$ is a valid mixer for $B$ satisfying the conditions of Equations (5) and (6), see also Figure 2.*

**Proof.** Any $|v\rangle \in B$ is in the null space of $H_{M,C}$, i.e., $H_{M,C}|v\rangle = 0$ and hence $U_{M,C}|v\rangle = I$. Therefore, $U_{M,B} U_{M,C} |v\rangle = U_{M,B} |v\rangle \in B$, and $U_{M,C} U_{M,B} |v\rangle = U_{M,C} |w\rangle = |w\rangle$ with $|w\rangle \in B$ which means the feasible subspace is preserved. Condition (6) follows similarly form the fact that $U_{M,C} |v\rangle = I$ for any $|v\rangle \in B$. \hfill $\square$

Corollary 2 naturally holds as well for any linear combination of mixers, i.e., $H_{M,B} + \sum q_i H_{M,C}$, is a mixer for the feasible subspace $Sp(B)$ as long as $Sp(C_i) \cap Sp(B) = \emptyset$, $\forall i$. At first, it might sound counter intuitive that adding more terms to the mixer results in more efficient decomposition into basis gates. However, as we will see in 3.4 it can lead to cancellations due to symmetry considerations.

Next, we describe feasible initial states which are given by eigenvectors of $U_M$.

**Corollary 3** (Feasible initial states.). *Given the setting in Theorem 1 with a symmetric transition matrix $T$. Let $(\lambda, v)$ an eigepair of $T$, then $(\lambda, Ev)$ is an eigepair of $H_M$ and $(e^{-i\lambda}, Ev)$ an eigepair of $U_M$, where $E = \{|x_j\rangle\}_{j \in J}$ as defined in Equation (12). Since feasible initial states need to be unit vectors,

$$
|\phi_0\rangle = \frac{1}{\|v\|_2} Ev
$$

is a feasible initial state, if $v$ is an eigenvector of $T$.***

**Proof.** Let $(\lambda, v)$ an eigepair of $T$. Then, $H_M Ev = ETE^T Ev = ET v = \lambda Ev$, so $(\lambda, Ev)$ is an eigenpair of $H_M$. The connection between $H_M$ and $U_M$ is general knowledge from linear algebra. Finally, we have that $\|Ev\|_2 = \|\sum_{j \in J} |x_j\rangle v_j\|_2 = \sum_{j \in J} \sqrt{|x_j|} |v_j|^2 = \sum_{j \in J} |v_j|^2 = \|v\|_2$. \hfill $\square$
Figure 2: Corollary 2 shows that adding a mixer with support outside $\text{Sp}(B)$ is also a valid mixer for $B$.

An example illustrating Corollary 3 is provided by the transition matrix $T \in \mathbb{R}^{4 \times 4}$ with zero diagonal and all other entries equal to one. A unit eigenvector of $T$, which fulfills Theorem 1, is $v = \frac{1}{2}(1, 1, 1, 1)^T$.

For any $B = \{ |z_1 \rangle, |z_2 \rangle, |z_3 \rangle, |z_4 \rangle \}$ a feasible initial state is given by the uniform superpositions of these states, since

$$
\frac{1}{\|v\|_2} Ev = \frac{1}{2} (|z_1 \rangle + |z_2 \rangle + |z_3 \rangle + |z_4 \rangle)(1, 1, 1, 1)^T.
$$

This result holds irrespective of what the states are and which dimension they have.

**Theorem 2** (Products of mixers for subspaces). Given the same setting as in Theorem 1. For any decomposition of $T$ into a sum of $Q$ symmetric matrices $T_q$, in the following sense

$$
T = \sum_{q=1}^{Q} T_q, \quad (T_q)_{i,j} = \begin{cases} 
\text{either} & (T)_{ij}, \\
\text{or} & 0,
\end{cases}
$$

we construct the mixing operator via

$$
U_M(\beta) = \prod_{n=1}^{N} e^{-i\beta T_m}.
$$

If all entries of $T$ are positive, then $U_M$ provides transitions between all pairs of feasible states, i.e. condition (6) is fulfilled, if for all $1 \leq j, k \leq |J|$ there exist $r_m \in \mathbb{N} \cup \{0\}$ (possibly depending on the pair) such that

$$
\left( \prod_{m=1}^{M} T_m^{r_m} \right)_{j,k} \neq 0.
$$

**Proof.** Combining Equations (13) and (14) we have that

$$
\langle x_j | U_M(\beta) | x_k \rangle = \sum_{j_1=0, j_2=0, \ldots, j_M=0}^{\infty} \frac{(-i)^{j_1+j_2+\cdots+j_M} (T_{q_1}^{j_1} T_{q_2}^{j_2} \cdots T_{q_M}^{j_M})_{j,k}}{j_1! j_2! \cdots j_M!}
$$

$$
= \sum_{j=1}^{\infty} \frac{(-i)^{j}}{j!} \sum_{j_1, \ldots, j_M \text{ s.t. } \sum_{m=1}^{M} j_m = j} (T_{q_1}^{j_1} T_{q_2}^{j_2} \cdots T_{q_M}^{j_M})_{j,k}
$$

Using that $T$ only as positive entries and the condition in Equation (19) the same argument as in Theorem 1 can be used to show that $U_M(\beta)$ is not the zero function and therefore we have transitions between all pairs of feasible states.

As Theorem 1 leaves a lot of freedom for choosing valid transition matrices we will continue by describing important examples for $T$.
Figure 3: Examples of overlap between two states for the case $|B| = 4$. The overlap is independent of what the states in $B = \{|z_0\rangle, |z_1\rangle, |z_2\rangle, |z_3\rangle\}$ are. The comparison for different $T$ show that there exists a $\beta$ such that the overlap is nonzero, except for $T_{2^+3}$ which, as expected, does not provide transitions between $|z_0\rangle$ and $|z_3\rangle$.

### 3.2 Transition matrices for mixers

Theorem 1 provides conditions for the construction of mixer Hamiltonians that preserve the feasible subspace and provide transitions between all pairs of feasible computational basis states, namely

1. $T \in \mathbb{R}^{[J] \times [J]}$ is symmetric, and
2. for all $1 \leq j, k \leq |J|$ there exist an $r_{j,k} \in \mathbb{N} \cup \{0\}$ such that $(T^r)_{j,k} \neq 0$.

Remarkably, these conditions depend only on the dimension of the feasible subspace $|J| = \dim(\text{Sp}(B)) = |B|$, but are independent of the specific states that $B$ consists of. In addition, Corollary 1 shows that these conditions are robust with respect to reordering of rows if in addition columns are reordered in the same way. Moreover, Equation (15) shows that also the overlap between computational basis states $|x_j\rangle, |x_k\rangle \in B$ is independent of the specific states that $B$ consists of and only depends on $T$, since the right hand side of the expression

$$\langle x_j \mid U_M(t) \mid x_k \rangle = \sum_{m=0}^{\infty} \frac{(-it)^m}{m!} (T^m)_{j,k}, \quad (21)$$

is independent of elements in $B$. This allows to describe and analyze valid transition matrices by only knowing the number of feasible states, i.e., $|B|$. What these specific states are is irrelevant, unless one wants to look at what an optimal mixer is, which we will come back to in Section 3.4. Figure 3 provides a comparison of some mixers described in the following with respect to the overlap between different states.

In the following, we denote the matrix for pairs of indices whose binary representation have Hamming distance equal to $d$ as

$$T_{\text{Ham}(d)}, \quad \text{with } (T_{\text{Ham}(d)})_{i,j} = \begin{cases} 1, & \text{if } d_{\text{Hamming}}(\text{bin}(i), \text{bin}(j)) = d, \\ 0, & \text{else,} \end{cases} \quad (22)$$
Examples of the structure of $T_{\text{Ham}(d)}$ can be found in Figure 4. Furthermore, it will be useful to denote the matrix which has two none-zero entries at $(k, l)$ and $(l, k)$ as $T_{k \leftrightarrow l}$, with $(T_{k \leftrightarrow l})_{i,j} = \begin{cases} 1, & \text{if } (i, j) = (k, l) \text{ or } (i, j) = (l, k), \\ 0, & \text{else}. \end{cases}$ (23)

Before we start, we point out that the diagonal entries of $T$ can be chosen to be zero, because $|T_{0,j,j}| = 1 \neq 0$ for all $j \in J$. Although trivial, we will repeatedly use that $v = \frac{1}{\sqrt{|J|}}(1, 1, \cdots, 1)^T$ is an eigenvector of a matrix $F \in \mathbb{C}^{|J| \times |J|}$ if the sum of all rows are a multiple of $v$.

### 3.2.1 Hamming distance one mixer $T_{\text{Ham}(1)}$

The matrix $T_{\text{Ham}(1)} \in \mathbb{R}^{|J| \times |J|}$ fulfills Theorem 1 when $|J| = 2^n, n \in \mathbb{N}$. The symmetry of $T_{\text{Ham}(1)}$ is due to the fact that the Hamming distance is a symmetric function. Using the identity

$$T_{\text{Ham}(k)} T_{\text{Ham}(1)} = T_{\text{Ham}(1)} T_{\text{Ham}(k)} = (n - (k - 1)) T_{\text{Ham}(k-1)} + (k + 1) T_{\text{Ham}(k+1)}$$

(24)

it can be shown that

$$T_{\text{Ham}(1)}^k = \sum_{j=1}^{k-1} c_k T_{\text{Ham}(j)} + k! T_{\text{Ham}(k)},$$

(25)

where $c_k$ are real coefficients. Therefore, it is clear that $T_{\text{Ham}(1)}^k$ reaches all states with Hamming distance $K$. Furthermore, $v = \frac{1}{\sqrt{|J|}}(1, 1, \cdots, 1)^T$ is a unit eigenvector of $T_{\text{Ham}(1)}$ since the sum of each row is $n$. This is because there are exactly $n$ other states with Hamming distance one for each bitstring.

### 3.2.2 All-to-all mixer $T_A$

We denote the matrix with all but the diagonal entries equal to one as

$$T_A, \quad \text{with } (T_A)_{i,j} = \begin{cases} 1, & \text{if } i \neq j, \\ 0, & \text{else}. \end{cases}$$

(26)

Trivially, $T_A \in \mathbb{R}^{|J| \times |J|}$ fulfills Theorem 1 and $v = \frac{1}{\sqrt{|J|}}(1, 1, \cdots, 1)^T$ is a unit eigenvector of $T_A$ since the sum of each row is $|J| - 1$. 

8
Figure 5: Valid (white) and invalid (black) transitions between pairs of states, as defined in Theorem 2 for Trotterized mixer Hamiltonians. The first row shows that for $T_1 = T_{O(1),c}$ and $T_2 = T_{E(1)}$, the mixer $U = e^{-i\beta T_1} e^{-i\beta T_2}$ does not provide transitions between all pairs of feasible states, although $U = e^{-i\beta (T_1 + T_2)}$ does.

3.2.3 (Cyclic) Nearest int mixer $T_{\Delta}/T_{\Delta,c}$

Inspired by the stencil of finite-difference methods we introduce $T_{\Delta}$, $T_{\Delta,c} \in \mathbb{R}^{|J| \times |J|}$ as the matrices with off-diagonal equal to one

$$T_{\Delta}, \quad \text{with } (T_{\Delta})_{i,j} = \begin{cases} 1, & \text{if } i = j + 1 \lor i = j - 1, \\ 0, & \text{else}, \end{cases}$$

$$T_{\Delta,c}, \quad \text{with } (T_{\Delta,c})_{i,j} = \begin{cases} 1, & \text{if } i = (j + 1) \mod n \lor i = (j - 1) \mod n, \\ 0, & \text{else}. \end{cases} \quad (27)$$

Both matrices fulfill Theorem 1. Symmetry holds by definition and it is easy to see that the k-th off-diagonal of $T_{\Delta}$ and $T_{\Delta,c}$ is nonzero for $1 \leq k \leq |J|$.

For the nearest integer mixer $T_{\Delta}$ it is known that

$$v_k = (\sin(c), \sin(2c), \cdots, \sin(|J|c)), \quad c = \frac{k\pi}{|J| + 1} \quad (28)$$

are eigenvectors for $1 \leq k \leq |J|$. For the cyclic nearest integer mixer, we have that the sum of each row/column of $T_{\Delta,c}$ is equal to two (except for $n = 1$ when it is one). Therefore, $v = \frac{1}{\sqrt{|J|}}(1, 1, \cdots, 1)^T$ is a unit eigenvector.

3.2.4 Products of mixers and $T_E, T_O$

In some cases, it will be necessary to use Theorem 2 to implement mixer unitaries. When splitting transitions matrices into odd and even entries the following definition is useful. Let the matrix with entries in the d-th off-diagonal for even rows equal to one

$$T_{E(d)}, \quad \text{with } (T_{E(d)})_{i,j} = \begin{cases} 1, & \text{if } i = j + d \lor i = j - d, \text{ and } i \text{ even,} \\ 0, & \text{else}, \end{cases} \quad (29)$$
and accordingly \( T_{O(d)} \) for odd rows. In addition, we will need \( T_{O(1),c} \) to be the cyclic version in the same way as in Equation (27). As an example, this allows to decompose \( T_{\Delta,c} = T_1 + T_2 \in \mathbb{R}^{n \times n} \) with \( T_1 = T_{O(1)} + T_{O(n-1)} = T_{O(1),c} \) and \( T_2 = T_{E(1)} \).

### 3.2.5 Random mixer \( T_{\text{rand}} \)

Finally, the upper triangular entries of the mixer \( T_{\text{rand}} \) are drawn from a continuous uniform distribution on the interval \([0, 1]\), and the lower triangular entries are chosen such that \( T \) becomes symmetric. Since the probability of getting a zero entry is zero, such a random mixer fulfills Theorem 1 with probability 1.

### 3.3 Decomposition of (constraint) mixers into basis gates

Given a set of feasible (computational basis) states \( B = \{ |x_j\rangle, \ j \in J, \ x_j \in \{0, 1\}^n \} \), we can use Theorem 1 to define a suitable mixer Hamiltonian. The next question is how to (efficiently) decompose the resulting mixer into basis gates. In order to do so we first decompose the Hamiltonian \( H_M \) into a weighted sum of Pauli-strings. A Pauli-string \( P \) is a Hermitian operator of the form \( P = P_1 \otimes \cdots \otimes P_n \) where \( P_i \in \{I, X, Y, Z\} \). Pauli-strings form a basis of the real vector space of all \( n \)-qubit Hermitian operators. Therefore, we can write

\[
H_M = \sum_{i_1, \ldots, i_n=1}^{4} c_{i_1, \ldots, i_n} \sigma_{i_1} \otimes \cdots \otimes \sigma_{i_n}, \quad c_{i_1, \ldots, i_n} \in \mathbb{R},
\]

with real coefficients \( c_{i_1, \ldots, i_n} \), where \( \sigma_1 = I, \sigma_2 = X, \sigma_3 = Y, \sigma_4 = Z \). After using a standard Trotterization scheme (which is exact for commuting Paul-strings),

\[
U_M(t) = e^{-itH_M} \approx \prod_{i_1, \ldots, i_n=1}^{4} e^{-itc_{i_1, \ldots, i_n} \sigma_{i_1} \otimes \cdots \otimes \sigma_{i_n}},
\]

it is well-established how to implement each of the terms of the product using basis gates, see Equation (32). We will discuss the effects of Trotterization in more detail in Section 3.5, as there are several important aspects to consider for a valid mixer.

\[
e^{-itP} = U \quad \text{Rz}(2t) \quad U^T \Rightarrow \begin{array}{c}
\begin{array}{c}
\begin{array}{c}
\begin{array}{c}
\vdots
\end{array}
\end{array}
\end{array}
\end{array}
\end{array}
\]

The standard way to compute the coefficients \( c_{i_1, \ldots, i_n} \) is given in Algorithm 1. For \( n \) qubits this requires to compute \( 4^n \) coefficients, as well as multiplication of \( 2^n \times 2^n \) matrices. However, most of these terms are expected to vanish. We therefore describe an alternative way to produce this decomposition, using the language of quantum chemistry. In the following we use the ladder operators used in the creation and annihilation operators from the second quantization formulation in quantum chemistry defined by

\[
a^\dagger = \frac{1}{2}(X - iY), \quad a = \frac{1}{2}(X + iY).
\]

\[
\begin{array}{c}
U_i = \begin{cases}
H, & \text{if } P_i = X, \\
SH, & \text{if } P_i = Y, \\
I, & \text{if } P_i = Z,
\end{cases}
\quad
(U^\dagger)_i = \begin{cases}
H, & \text{if } P_i = X, \\
HS^\dagger, & \text{if } P_i = Y, \\
I, & \text{if } P_i = Z.
\end{cases}
\end{array}
\]
Algorithm 1: Decompose $H_M$ given by Equation (8) into Pauli-strings via trace

| Data: Feasible states $B$ and transition matrix $T$ fulfilling Theorem 1 |
| Result: Coefficients $c_{i_1,...,i_n}$ of Pauli-strings, Equation (30) |

for $i_1 = 1,\ldots,4$ do
  ...
  for $i_n = 1,\ldots,4$ do
    $c_{i_1,...,i_n} = \frac{1}{2} \text{tr} (\sigma_{i_1} \otimes \cdots \otimes \sigma_{i_n} H_M)$
  end
end

Since $a |0\rangle = 0, a |1\rangle = |0\rangle$, where 0 is the zero vector, we have that $|0\rangle \langle 1| = a$. Since $a^\dagger |0\rangle = |1\rangle, a^\dagger |1\rangle = 0$, we have that $|1\rangle \langle 0| = a^\dagger$, and finally $a^\dagger a |0\rangle = 0, a^\dagger a |1\rangle = |1\rangle, a a^\dagger |0\rangle = |0\rangle, a a^\dagger |1\rangle = 0$, means that $|0\rangle \langle 0| = aa^\dagger$ and $|1\rangle \langle 1| = a^\dagger a$. Note that

$$a^\dagger a = \frac{1}{2} (I - Z), \quad aa^\dagger = \frac{1}{2} (I + Z). \quad (34)$$

As an example, consider the matrix $M = |01\rangle \langle 0| = |0\rangle \langle 1| \otimes |1\rangle \langle 0|$ which can be expressed with ladder operators as $M = a_1 a_2^\dagger$. Another example is given by $M = |01\rangle \langle 11| = a_1 a_2^\dagger a_2$. This approach clearly extends to the general case and leads to the Algorithm 2.

Algorithm 2: Decompose $H_M$ given by Equation (8) into Pauli-strings directly

| Data: Feasible states $B$ and transition matrix $T$ fulfilling Theorem 1 |
| Result: Non-vanishing coefficients $c_{i_1,...,i_n}$ of Pauli-strings, Equation (30) |

S=0
for all $1 \leq j, k \leq |J|, s.t. (T)_{j,k} \neq 0$ do
  Define $x_L = x_J, x_R = x_{J_k}$
  for $1 \leq i \leq n$ do
    $(P_{j,k})_i = \begin{cases} 
    \frac{1}{2}(X + iY), & \text{if } (x_L)_i = 0 \land (x_R)_i = 1, \\
    \frac{1}{2}(X - iY), & \text{if } (x_L)_i = 1 \land (x_R)_i = 0, \\
    \frac{1}{2}(I + Z), & \text{if } (x_L)_i = 0 \land (x_R)_i = 0, \\
    \frac{1}{2}(I - Z), & \text{if } (x_L)_i = 1 \land (x_R)_i = 1.
    \end{cases}$
  end
  $S = S + (T)_{j,k} P_{j,k}$
end
simplify $S$ (e.g., using a library for symbolic mathematics)
this defines the non-vanishing coefficients $c_{i_1,...,i_n}$

A comparison of the complexity of the two algorithms is given in Table 1. The naive algorithm needs to perform a matrix-matrix-multiplication with matrices of the size $2^n \times 2^n$ for each of the $4^n$ coefficients. This quickly becomes prohibitive for larger $n$. The algorithm based on ladder operators requires resources that scale with the number of non-zero entries of the transition matrix $T$, which is much more favourable. In the end a symbolic mathematics library is used to simplify the expressions in order to create the list of non-zero Pauli-strings.

3.4 Optimality of mixers

On current NISQ devices, the noise level of two-qubit gate (CX) times and error rates are one order of magnitude higher than for single qubit gates ($U_3$). In addition, most devices lack all-to-all connectivity.
CX gates between these require SWAP operations, which consist of additional CX gates. An optimal mixer will therefore contain as few CX gates as possible. Since Pauli-strings are implemented according to Equation (32) we define the cost to implement $e^{-itH_M}$ as

$$\text{Cost}(H_M) = 2\sum_{i_1, \ldots, i_n | |c_{i_1, \ldots, i_n}| > 0, \text{len}(\sigma_{i_1} \otimes \cdots \otimes \sigma_{i_n}) > 1} (\text{len}(\sigma_{i_1} \otimes \cdots \otimes \sigma_{i_n}) - 1),$$

(35)

where we defined the length of a Pauli-string $P$ to be the number of literals that are not the identity. For instance $P = IXIY = I_X I_X I_Y I_Y = X_I Y_I$ has $\text{len}(P) = 2$. The Cost($H_M$) specifies the number of CX gates that are required to implement the mixer. A lower cost means fewer and/or shorter Pauli-strings. There are four interconnected factors that influence the cost to implement the mixer for a given $B$.

### 3.4.1 Transition matrix $T$

The larger $|B|$ the more freedom we have in choosing the transition matrix $T$ that fulfills Theorem 1. The combination of $T$ and the specific states of $B$ define the cost of the Hamiltonian. Unless one can find a way to utilize the structure of the states of $B$ to efficiently compute an optimal $T$, we expect this problem to be NP-hard. In practice, a careful analysis of the specific states of $B$ is required to determine $T$ such that the cost becomes low. We will revisit optimality for both unrestricted and restricted mixers with

### 3.4.2 Feasible initial state

Another aspect is the following. The cost to implement a mixer $T$ needs to be balanced with how costly it is to prepare a feasible initial state. This depends on the actual states in $B$ as well as on the number of layers in QAOA. However, Corollary 3 shows that the eigenvalues of $T$ determine the eigenvalues of $U_M$, which helps to more easily decide this question.

### 3.4.3 Adding mixers

Corollary 2 allows to add mixers with a kernel that contains Sp($B$). In general, also this is a combinatorial optimization problem which we do not expect to solve exactly with an efficient algorithm. However, we will provide a heuristic that can be used to reduce the cost of mixers in certain cases. We will provide more details in Section 5 where we discuss constrained mixers on some examples in detail.

### 3.4.4 Non-commuting Pauli-strings

Depending on the mixer – which depends on the transition matrix and addition of mixers outside the feasible subspace – one can influence the commutativity pattern of the resulting Pauli-strings. This is an intricate topic, which we discuss next.

### 3.5 Trotterizations

Algorithms 1 and 2 produce a weighted sum of Pauli-strings equal to the mixer Hamiltonian $H_M$ defined in Theorem 1. A further complication arises when the non-vanishing Pauli-strings of the mixer Hamiltonian $H_M$ do not all commute. In that case one can not realize $U_M$ exactly but has to find a suitable approximation/Trotterization, see Equation (31). Two Pauli-strings commute, i.e., $[P_A, P_B] = P_A P_B - P_B P_A = 0$ if and only if they fail to commute on an even number of indices $4$. An example is given in Figure 6.

This problem is similar to a problem for observables; how does one divide the Pauli-strings into groups of commuting families $4, 5$ such as to maximize efficiency and increase accuracy? In order to minimize the number of measurements required to estimate a given observable one wants to find a ”min-commuting-partition”; given a set of Pauli-strings from a Hamiltonian, one seeks to partition the strings
into commuting families such that the total number of partitions is minimized. This problem is NP-hard in general \cite{2}. However, based on Theorem 3 we expect our problem to be much more tractable.

For our case, it turns out that not all Trotterizations are suitable as mixing operators: they can either fail to preserve the feasible subspace, i.e., Equation (6), or fail to provide transitions between all pairs of feasible states, i.e., Equation (6). An example is given by \( B = \{001, 101, 110\} \) with the mixer \( H_M = \frac{1}{4}(XIX + YIY) + \frac{1}{4}(XXI + YYI) \) associated with \( T_S = T_{1+2} + T_{2+3}, \) see Section 5.1. Looking at Figure 6, these terms can be grouped into commuting families in two ways, which represent two (of many) different ways to realize the mixer unitary with basis gates.

1. The first possible Trotterization is given by \( U_1(\beta) = e^{-i\beta(XXI+IXX)} \) and \( U_2(\beta) = e^{-i\beta(YYI+IYY)} \). However, it turns out that \( \exists \beta \in \mathbb{R} \) such that \( |\langle 111 | U_1(\beta)U_2(\beta) | z \rangle| > 0 \) for all \( |z\rangle \in B \). This means that this Trotterization does not preserve the feasible subspace and does not represent a valid mixer Hamiltonian. The underlying reason for this is that the terms \( XXI \) and \( YYI \) are generated from the entry \( T_{1+2} \), but are split in this Trotterization. The same holds true for \( IXX \) and \( IYY \), which are generated via \( T_{2+3} \).

2. The second possible Trotterization is given by \( U_1(\beta) = e^{-i\beta(XXI+YYI)} \) and \( U_2(\beta) = e^{-i\beta(XXI+YYI)} \), which splits terms with respect to \( T_{1+2} \) and \( T_{2+3} \). In this case, we have that \( |\langle 100 | U_1(\beta)U_2(\beta) | 001 \rangle| = 0 \), so it does not provide an overlap between all feasible computational basis states. This can be understood via Theorem 2. We have that \( (T_{1+2} T_{2+3})_{3,1} = 0 \) for all \( n_1, n_2 \in \mathbb{N} \), so one can not "reach" \( |100 \rangle \) from \( |001 \rangle \). The opposite is not true: we have that \( (T_{1+2} T_{2+3})_{1,3} = 1 \), so \( \exists \beta \) such that \( |\langle 001 | U_1(\beta)U_2(\beta) | 100 \rangle| > 0 \).

We have just learned that it is a bad idea to Trotterize terms that belong to a non-zero entry of \( T \), i.e. to \( T_{j+3} \). Therefore, we need to exclude that there are cases where not all non-vanishing Pauli-strings of \(|x_j\rangle \langle x_i| + |x_i\rangle \langle x_j| \) commute. If that were the case, it might not be possible to realize a mixer for a certain feasible subspace. Luckily, the following Theorem shows that it is always possible to realize a mixer by Trotterizing according to non-zero entries of \( T = \sum_{j \in \mathbb{J} \subset \mathbb{J}} T_{j+3} \).

**Theorem 3 (Pauli strings for \( T_{j+3} \) commute).** Let \(|z\rangle \langle w| \) be two computational basis states in \( \mathbb{C}^2^n \). Then all non-vanishing Pauli-strings \( \sigma_{i_1} \otimes \cdots \otimes \sigma_{i_n} \) of the decomposition

\[
|z\rangle \langle w| = \sum_{i_1, \ldots, i_n = 1}^n c_{i_1, \ldots, i_n} \sigma_{i_1} \otimes \cdots \otimes \sigma_{i_n}, \quad c_{i_1, \ldots, i_n} \in \mathbb{R}, \quad (36)
\]

commute.

**Proof.** We will proof the following more general assertion by induction. Let \( P_{1+1}, P_{1+2} \) two non-vanishing Pauli-strings of the decomposition of \(|z\rangle \langle w| + |w\rangle \langle z| \), and \( P_{1-1}, P_{1-2} \) two non-vanishing Pauli-strings of the decomposition of \( i(|z\rangle \langle w| - |w\rangle \langle z|) \). Then \( [P_{1+1}, P_{1+2}] = 0, [P_{1-1}, P_{1-2}] = 0 \) and \([P_{1+1}, P_{1-2}] \neq 0\). We will use that two Pauli-strings commute if and only if they fail to commute on an even number of indices \[4\].

For \( n = 1 \) we have the following cases.

\[
A_1 = \frac{1}{\sqrt{2}}(|z\rangle + |w\rangle), \quad B_1 = \frac{1}{\sqrt{2}}(|z\rangle - |w\rangle) \Rightarrow \begin{cases} 
I + Z, \quad &\text{if} \; (z,w) = (0,0), \\
X, \quad &\text{if} \; (z,w) = (0,1), \\
X, \quad &\text{if} \; (z,w) = (1,0), \\
I - Z, \quad &\text{if} \; (z,w) = (1,1),
\end{cases}
\]

\[
C_1 = \frac{1}{\sqrt{2}}(|z\rangle - |w\rangle), \quad D_1 = \frac{1}{\sqrt{2}}(|z\rangle + |w\rangle) \Rightarrow \begin{cases} 
0, \quad &\text{if} \; (z,w) = (0,0), \\
-Y, \quad &\text{if} \; (z,w) = (0,1), \\
+Y, \quad &\text{if} \; (z,w) = (1,0), \\
0, \quad &\text{if} \; (z,w) = (1,1).
\end{cases}
\]

It is trivially true that \( [P_{1+1}, P_{1+2}] = 0, [P_{1-1}, P_{1-2}] = 0 \) since the maximum number of Pauli-strings is two, and in that case one of the Pauli-strings is the identity. Moreover, \( P_{1-} \) is nonzero only when \( z \neq w \). In that case \( [P_{1+1}, P_{1-2}] = [X, \pm Y] \neq 0 \).
n → n + 1. We assume the assumptions hold for two computational basis states \(|z\rangle, |w\rangle \in \mathbb{C}^2\). Then there are the following four cases

\[
A_{n+1} = \frac{1}{2} (|wx\rangle \langle zy| + |wx\rangle \langle zy|) - \frac{1}{2} (|wy\rangle \langle zy| - |wy\rangle \langle zy|)
\]

where \(A_n = (|z\rangle \langle w| + |w\rangle \langle z|), B_n = i(|z\rangle \langle w| - |w\rangle \langle z|)\).

**Case x = y.** According to our assumptions that all non-vanishing Pauli-strings for \(A_n\) commute, the same holds for \(A_{n+1} = A_n \otimes (I \pm Z)\). Since \(B_{n+1} = 0\) the rest of the assertions are trivially true as there are no non-vanishing Pauli-strings.

**Case x ≠ y.** Our assumptions mean that non-vanishing Pauli-strings of \(A_n\) fail commute on an odd number of indices with non-vanishing Pauli-strings of \(B_n\). Therefore, non-vanishing Pauli-strings of \(A_{n+1} = A_n \otimes X \pm B_n \otimes Y\) fail to commute on an even number of indices, and, hence, commute. The same argument holds for \(B_{n+1} = B_n \otimes X \pm A_n \otimes Y\). Finally, we prove that non-vanishing Pauli-strings of \(A_{n+1}\) and \(B_{n+1}\) do not commute. Either Pauli strings \(P_{1 \pm r}\) and \(P_{1 \pm s}\), stem from \(A_n \otimes X \oplus B_n \otimes X\), respectively, or they stem from \(±A_n \otimes Y\) and \(±B_n \otimes Y\), respectively. In both cases, the number of commuting terms does not change, so non-vanishing Pauli-strings of \(A_{n+1}\) and \(B_{n+1}\) do not commute.

The proof in Theorem 3 inspires the following algorithm to decompose \(H_M\) into Pauli-strings. For each item in the list \(S\) that the algorithm produces, all Pauli-strings commute.

**Algorithm 3:** Decompose \(H_M\) given by Equation (8) into Pauli-strings directly

**Data:** Feasible states \(B\), transition matrix \(T\) (zero diagonal) fulfilling Theorem 1

**Result:** Non-vanishing coefficients \(c_{i_1, \ldots, i_n}\) of Pauli-strings, Equation (38)

\[
S = [\ , \ ]
\]

for all \(1 \leq j < k \leq |J|\), s.t. \((T)_{j,k} \neq 0\)

- Recursively compute \(A_n\) and \(B_n\) for \(|w\rangle = |x_{J_j}\rangle\) and \(|z\rangle = x_{J_k}\rangle\) via Equation (38)

end

check for cancellations and return \(S\)

We can illustrate the difference between Algorithms 2 and 3 for \(B = \{|01\rangle, |10\rangle\}\) and \(T_{1 \leftrightarrow 2}\). With Algorithm 2 we have \(P_{1,2} = \frac{1}{2}(X - iY)(X + iY)\) and \(P_{2,1} = \frac{1}{2}(X + iY)(X - iY)\) which can be simplified to \(S = P_{1,2} + P_{2,1} = \frac{1}{2}(XX + YY)\). With Algorithm 3 we have \(A_1 = X, B_1 = Y\) and \(S = A_2 = \frac{1}{2}(A_1X + B_1Y) = \frac{1}{2}(XX + YY)\) without the need to simplify the expression.

As shown above, Trotterizations can also lead to missing transitions. It is suggested in [6] that it is useful to repeat mixers within one mixing step, which corresponds to \(r > 1\) in Equation (6). However, as we see in Figure 3 there can be more efficient ways to get mixers which provide transitions between all pairs of feasible states. One way to do so is to construct an exact Trotterization (restricted to the feasible subspace) as described in [13]. However, the ultimate goal is **not** to avoid Trotter errors, but rather to provide transitions between all pairs of feasible states. We will revisit the topic of Trotterizations in Section 5 in more detail for each case and show that there are more efficient ways to do so.
Pauli-strings which contain exactly two

sentations. According to Section 3.2.1 a feasible initial state is given by uniform superposition

This assumes that the feasible states in

Hamiltonian is therefore given by

The last identity in Equation (39) shows that

Furthermore, using Equation (12) we have that

The Hamiltonian of the standard full mixer for n qubits can be written as

For this case

is the number of nonzero entries of \( T \).

4 Full/Unrestricted mixer

We start by applying the proposed algorithm to the case without constraints, i.e., for the case \( g = 0 \) in Equation [1], in order to check for consistency and new insight. We will see that the presented approach is able to reproduce the “standard” X mixer as one possibility, but provides a more general framework.

For this case \( B = \{ |x_j\rangle, j \in J, x_j \in \{0,1\}^n \}, J = \{ i, 1 \leq i \leq 2^n \} \) which means that \( Sp(B) = \mathcal{H} \).

Furthermore, using Equation (12) we have that \( H_{M,B} = T \), since \( E \) is the identity.

4.1 \( T_{\text{Ham}(1)} \) aka ”standard” full mixer

The Hamiltonian of the standard full mixer for n qubits can be written as

The last identity in Equation (39) shows that \( H_M \) is created by the transition matrix given by \( T_{\text{Ham}(1)} \). This assumes that the feasible states in \( B \) are ordered from the smallest to the largest integer representation. According to Section [3.2.1] a feasible initial state is given by uniform superposition \( |\phi_0\rangle = \frac{1}{\sqrt{2^n}} \sum_{j \in J} |x_j\rangle = \frac{1}{\sqrt{2^n}} (1,1,\cdots,1)^T = |+\rangle^\otimes n \), which is a well established fact.

4.2 All-to-all full mixer

For \( |J| = 2^n \) the full mixer \( T_A \) can be written as \( T_A = \sum_{j=1}^n T_{\text{Ham}(j)} \). For the case \( T_{\text{Ham}(2)} \) the resulting Hamiltonian \( H_M \) does not provide transitions between all pairs of feasible states, but we observe that \( H_M = \sum_{j,k \in J} (T_{\text{Ham}(2)})_{j,k} |x_j\rangle \langle x_k| = \sum_{j_1} \sum_{j_2=j_1+1}^n X_{j_1} X_{j_2} \), i.e., \( H_M \) consists of all \( \binom{n}{2} \) possible pairs of Pauli-strings which contain exactly two \( X \)'s. For \( m \leq n \) this can be further generalized to

which consists of all \( \binom{n}{m} \) possible pairs of Pauli-strings with exactly \( m \) \( X \)'s. The resulting mixer Hamiltonian is therefore given by

This means that the mixer consists of the standard mixer plus \( \binom{n}{k} \) applications of Pauli \( X-k \) strings for \( k \) from 2 to \( n \), which is a large overhead with respect to the standard X-mixer. According to Section [3.2.2] a feasible initial state is given by uniform superposition \( |\phi_0\rangle = \frac{1}{\sqrt{2^n}} \sum_{j \in J} |x_j\rangle = \frac{1}{\sqrt{2^n}} (1,1,\cdots,1)^T = |+\rangle^\otimes n \), which is the same as for the standard mixer.

| Algorithm | \( \mathcal{O}(2^n) \) | \( \mathcal{O}(n\gamma) \) | \( \mathcal{O}(n\gamma) \) |
|-----------|-----------------|-----------------|-----------------|
| runtime   | \( \mathcal{O}(2^n) \) | \( \mathcal{O}(n\gamma) \) | \( \mathcal{O}(n\gamma) \) |
| memory    | \( \mathcal{O}(2^n) \) | \( \mathcal{O}(n\gamma) \) | \( \mathcal{O}(n\gamma) \) |

Table 1: Comparison of the complexity of the two algorithms for \( n \) qubits. Here, \( \gamma \) is the number of non-zero entries of \( T \).
requirements. The standard mixer is the only one which does not require CX gates and is the most
impression. Table 2 shows a dramatic difference between the different mixers with regards to resource
rest of the transition matrices the

\[ \text{Ham}(T_n) \]

\[ \#U_3 \]

\[ \#\text{CX} = \text{Cost}(H_M) \]

| \( n \) | 1 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| \( T_{\text{Ham}(1)} \) | 2 | 8 | 24 | 64 | 160 | 384 | 1 | 2 | 3 | 4 | 5 | 6 | 0 | 0 | 0 | 0 | 0 | 0 |
| \( T_A \) | 2 | 16 | 96 | 512 | 2560 | 12288 | 1 | 2 | 3 | 4 | 5 | 6 | 0 | 2 | 10 | 34 | 98 | 258 |
| \( T_{\Delta,c} \) | 2 | 12 | 28 | 60 | 124 | 252 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 2 | 12 | 44 | 132 | 356 |
| \( T_{\Delta} \) | 2 | 8 | 22 | 52 | 114 | 240 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 4 | 20 | 68 | 196 | 516 |
| \( T_{\text{rand}} \) | 2 | 16 | 96 | 512 | 2560 | 12288 | 2 | 4 | 6 | 8 | 10 | 12 | 0 | 10 | 86 | 552 | 3260 | 17650 |

Table 2: Full/unrestricted mixer case for \( n \) qubits, i.e., \( |B| = 2^n \). Comparison of the total Hamming
distance of the transition matrix \( T \) as well as resulting requirements for implementations in terms of
single- and two-qubit gates for different \( T \).

### 4.3 (Cyclic) nearest integer full mixer

The resulting mixer for \( T_{\Delta}/T_{\Delta,c} \) involve exponentially many Pauli-strings with increasing \( n \). The following
shows \( T_{\Delta} \) for \( 1 \leq n \leq 4 \).

\[
H_M^{n=1} = X_1, \quad H_M^{n=2} = I \otimes H_M^{n=1} + \frac{1}{2} (X_1 X_2 + Y_1 Y_2), \quad H_M^{n=3} = I \otimes H_M^{n=2} + \frac{1}{4} (X_1 X_2 X_3 - X_1 Y_2 Y_3 + Y_1 X_2 Y_3 + Y_1 Y_2 X_3), \quad H_M^{n=4} = I \otimes H_M^{n=3} + \frac{1}{8} (X_1 X_2 X_3 X_4 - X_1 X_2 Y_3 Y_4 - X_1 Y_2 X_3 Y_4 - X_1 Y_2 Y_3 X_4 \\
+ Y_1 X_2 X_3 X_4 + Y_1 X_2 Y_3 X_4 + Y_1 Y_2 X_3 X_4 + Y_1 Y_2 Y_3 Y_4). \]

According to Section 3.2.3 feasible initial states for \( T_{\Delta} \) are given by \( v_k/\|v_k\|^2 \) as defined in Equation (28).

For \( T_{\Delta,c} \) uniform superposition \( |\phi_0\rangle = \frac{1}{\sqrt{2^n}} \sum_{j \in J} |x_j\rangle = \frac{1}{\sqrt{2^n}} (1,1,\cdots,1)^T = |+\rangle^{\otimes n} \) is a feasible initial
state.

### 4.4 Comparison and optimality of full mixers

It would be convenient to have a condition on the transition matrix for the optimality of the resulting
mixer. We define the total Hamming distance of \( T \) to be

\[
\text{Ham}(T) = \sum_{j, k=1 \atop |(T)_{j, k}| > 0} |j| d_{\text{Hamming}}(\text{bin}(i), \text{bin}(j)).
\]

As a first instinct one might suspect that the mixer with minimal Hamming distance, also minimizes
the cost. However, this turns out to be false, because of cancellations when more terms in \( T \) are nonzero.
Table 2 gives a comparison of the total Hamming distance and cost for different full mixers. The
standard full mixer has total Hamming distance \( \text{Ham}(T) = (n - 1)2^n \), as there are \( 2^n \) states with each
\( n - 1 \) states that have Hamming distance 1. The all to all full mixer has \( \text{Ham}(T) = 2^n \sum_{k=1}^{n} k \binom{n}{k} \). The
rest of the transition matrices the \( \text{Ham}(T) \) is not that straightforward to calculate, but the table gives
an impression. Table 2 shows a dramatic difference between the different mixers with regards to resource
requirements. The standard mixer is the only one which does not require CX gates and is the most
efficient to implement. Furthermore, as the resulting Pauli terms for the full mixers given by \( T_{\text{Ham}(1)} \)
and \( T_A \) consist only of \( J \) and \( X \) and therefore commute they can be implemented without Trotterization.
For the mixers given by \( T_{\Delta}, T_{\Delta,c} \) and \( T_{\text{rand}} \) on the other hand, not all Pauli-strings commute which results
in the need for Trotterization. We continue with the case of constrained mixers.
5 Constrained mixers

We start by describing what is known as the "XY"-mixer [6, 13, 3], before we explore more general cases. Our framework provides additional insights into this case and inspires further improvement of above algorithms with respect to the optimality of the mixers, described in Section 3.4, by (possibly) reducing the length of Pauli-strings. For this case, we will analyze $T_{\Delta,c}$ only.

5.1 "One-hot" aka "XY"-mixer

We are concerned with the case given by all computational basis states with exactly one "1", i.e., $B = \{|x\rangle, x_j \in \{0,1\}^n, \text{s.t.} \sum x_j = 1\}$. These states are sometimes referred to as "one-hot". We have that $n = |B|$ is the number of qubits. After some motivating examples we present the general case for constructing mixers for any $n > 2$. Feasible initial states depend on $T$ and are given by Corollary 3. For example for $T = T_{\Delta,c}$ a feasible initial state is given by the $W_n$-state $|\phi_0\rangle = \frac{1}{\sqrt{n}} (|000...001\rangle + |000...010\rangle + ... + |100...000\rangle)$. (44)

An efficient algorithm for this with logarithmic (in $n$) time complexity is presented in [1]. For $T = T_{\Delta}$ a feasible initial state is given by

$$|\phi_0\rangle = \frac{1}{\sqrt{|v_n|}} \sin(c) |000...001\rangle + \sin(2c) |000...010\rangle + ... + \sin(nc) |100...000\rangle,$$  

(45)

where $c$ and $v_n$ are defined in Equation (28).

5.1.1 Case $n = 2$

The smallest, non-trivial case is given by $B = \{|01\rangle, |10\rangle\}$. For any $b \in \mathbb{R}, |b| > 0$ the transition matrix $T = \left( \begin{array}{cc} b & 0 \\ 0 & \frac{1}{b} \end{array} \right)$ fulfills Theorem 1 and leads to the mixer $H_M = \frac{b}{2}(XX + YY) + \frac{1}{2}(II - ZZ)$. Since we want to minimize Cost($H_M$) given in Equation (35), we set $d = 0$, which results in the Cost($H_M$) = 4. However, by using Corollary 2 there is room for further reducing the cost. We can add the mixer for $C = \{|00\rangle, |11\rangle\}$ since $B \cap C = \emptyset$. Using the same $T$ (setting $d = 0$) gives $H_{M,B} + H_{M,C} = bXX$, (46)

which has Cost($H_M$) = 2. No Trotterization is needed in this case.
5.1.2 Case \( n = 3 \)

We continue with \( B = \{001\}, \{010\}, \{100\} \}. For the transition matrix \( T = aT_{1\rightarrow 2} + bT_{2\rightarrow 3} + cT_{1\rightarrow 3} \), \( a, b, c \in \mathbb{R} \) this results in the mixer

\[
H_{M,B} = \frac{a}{4}(XX + YY)(I + Z) + \frac{b}{4}(I + Z)(XX + YY) + \frac{c}{4}(X(I + Z)X + Y(I + Z)Y),
\]

with associated cost \( \text{Cost}(H_M) = 24 + 12c \) for \( a = b = 1, c \in \{0, 1\} \). In this case Corollary 2 allows us to add the mixer for \( C = \{110\}, \{101\}, \{011\} \} since \( B \cap C = \emptyset \). The mixer

\[
H_M = H_{M,B} + H_{M,C} = \frac{a}{2}(XXI + YYI) + \frac{b}{2}(XIY + YIX) + \frac{c}{2}(IIX + IYY),
\]

has cost \( \text{Cost}(H_M) = 8 + 4c \) for \( a = b = 1, c \in \{0, 1\} \). However, this mixer can not be realized, since not all terms of \( H_M \) commute. Figure 6 shows two ways to put the graph into commuting Pauli-terms, with only one way to preserve the feasible subspace, as discussed in Section 3.3 A Trotterized mixer with the lowest

5.1.3 The general case \( n > 2 \)

We start with the observation that for any symmetric \( T \in \mathbb{R}^{n \times n} \) with zero diagonal we have

\[
H_{M,B} = \sum_{j=1}^{n} \sum_{k=j+1}^{n} (T)_{j,k} \tilde{P}^{j,k}_{l}, \quad (\tilde{P}^{j,k}_{l})_{l} = \begin{cases} 1, & \text{if } l \in \{j, k\}, \\ \frac{1}{2^n}, & \text{if } x_l = z_l = 0 \end{cases}
\]

The cost for implementing one of the entries, i.e., \( e^{-i\beta \tilde{P}^{j,k}_{l}} \) is given by the recursive formula

\[
\text{Cost}(\tilde{P}^{j,k}_{l}) = \sum_{l=2}^{n} 2(l-1)f_n^l, \quad n > 2, \quad f_n^l = f_{n-1}^l + f_{n-1}^{l-1}, \quad f_2^l = \begin{cases} 2, & \text{if } l = 2, \\ 0, & \text{else,} \end{cases}
\]

where \( f_n^l \) is Pascal's triangle starting with 2 instead of 1. Example for the resulting costs for different transition matrices can be seen in Table 3.

The cost of the mixers can be considerably reduced by adding mixers generalized from case \( n = 3 \). If the entries \( (T)_{i\rightarrow j} \) of \( T \) are non-zero, we can add mixers for each of the \( 2^n - 2 \) pairs of states \( x \in \{0, 1\}^n \) that fulfill that \( (x_i = 0 \land x_j = 1) \lor (x_i = 1 \land x_j = 0) \). We can enumerate them with \( 0 \leq l \leq 2^n - 2 \) by \( \tilde{B}^{i,j}_{l} = \{|x\}, x \in \{0, 1\}^n, \text{ s.t. } x_{i, i-j} = \text{bin}(l)\} \), where \( x_{i, i-j} \) removes the indices \( i \) and \( j \) of \( x \). We have that \( B \cap \tilde{B}^{i,j}_{l} = \emptyset \). We observe that for \( n \geq 2 \) let \( |x\}, |z\} \) with \( \text{Ham}(x, z) = 2 \), i.e., the strings \( x, z \) differ at exactly two positions we have that

\[
|x\} \langle z | + | z \} \langle x | = \begin{cases} (X + Y), & \text{if } x_l \neq z_l, \\ (I + Z), & \text{if } x_l = z_l = 0, \\ (I - Z), & \text{if } x_l = z_l = 1. \end{cases}
\]
Adding these mixers for each nonzero entries $T_{j \leftrightarrow k}$ of $T$ hast the effect to sum over all possible combinations of $(I \pm Z)^{\otimes 2^{n-2}}$ which is equal to the identity. Therefore, we get the mixer

$$H_{M,B} + \sum_{i,j \in J} \sum_{l=0}^{2^{n-2}} H_{M,B_{i,j}} = \sum_{j=1}^{n} \sum_{k=j+1}^{n} (T)_{j,k} P_{j,k}, \quad P_{j,k} = X_i X_j + Y_i Y_j,$$

which reduces the cost of one term to $\text{Cost}(P_{j,k}) = 4$.

### 5.1.4 Trotterizations

Not all Pauli-strings of the mixer in Equation (52) commute. This necessitates a suitable and efficient Trotterization. We will use Theorem 2 and Theorem 3 to identify valid Trotterized mixers. As pointed out in [13] when $n$ is a power of two one can realize a Trotterization which is exact in the feasible subspace $B$. Termed simultaneous complete-graph mixer, this involves all possible pairs $(i,j)$ corresponding to a certain Trotterization of mixer for $T_B$. We will see that there are more efficient mixers that provide transitions between all pairs of feasible states.

Another possibility is to Trotterize $T_{A,s}$ or $T_{A}$ according to odd and even entries as described in Section 3.2.4. This is what is termed parity-partitioned mixer in [13]. However, fewer and fewer feasible states can be reached as $n$ increases, as we have seen in Figure 5. Repeated applications ($r > 0$ in Equation (6)) are necessary and $r$ increases with increasing $n$. Figure 7 shows a comparison of different Trotterizations. As the cost of the mixer is dictated by the number of non-zero entries of the transition matrix, it is more efficient to add mixers for off-diagonals according to $\sum_{i \in I} (T_{O(i)} + T_{E(i)})$ for some suitable index set $I$.

### 5.2 General cases

In this section we analyze some specific cases that go beyond unrestricted mixers and mixers restricted to one-hot states.
Finally, we investigate the case

5.2.1 Example 1

We start by looking at the case $B = \{|100\rangle, |010\rangle, |011\rangle\}$. Using $T_\Delta = c_{1,2}T_{1+2} + c_{2,3}T_{2+3}$ and $T_{\Delta,c} = T_\Delta + c_{3,1}T_{3+1}$, this results in the mixer

$$H_{M,B} = c_{1,2} \frac{1}{4} (XX + YY)(I + Z) + c_{2,3} \frac{1}{4} (I + Z)(I - Z)X + c_{3,1} \frac{1}{4} (XXX + YXY + YYX - XYY)$$

with Cost($H_{M,B}$) = 12$c_{1,2} + 8c_{2,3} + 16c_{3,1}$. Here, ($c_{1,2}, c_{2,3}, c_{3,1}$) = (1, 1, 0) corresponds to $T_\Delta$ and (1, 1, 1) to $T_{\Delta,c}$. There is a lot of freedom adding mixers, which is summarized in Table 4. Adding more terms only increases the cost for this case. Overall, the most efficient mixers for $B$ given by

$$H_M = \frac{c_{1,2}}{2} \left[ XXI + XZX \text{ or } X XI + YYZ \right] + \frac{c_{2,3}}{2} \left[ (I + Z)IX \text{ or } I(I - Z)X \right] + \frac{c_{3,1}}{2} \left[ XXX - XYZ \text{ or } XXX + YXY \right]$$

with associated cost Cost($H_M$) = 6$c_{1,2} + 2c_{2,3} + 8c_{3,1}$. A valid Trotterization is given through splitting according to $T_{1+2}$.

5.2.2 Example 2

Finally, we investigate the case $B = \{|10010\rangle, |01110\rangle, |10011\rangle, |11101\rangle, |00110\rangle, |01010\rangle\}$, which restricts to 6 of the total $2^5 = 32$ computational basis states for 5 qubits. It is not a priori clear if for any (distinct) pair $T_{1,j_1}$ and $T_{j_2,j_2}$ all pairs of non-vanishing Pauli strings commute. In order to fulfill Equation 6 for $r = 1$, this means that one needs to Trotterize according to all pairs of $T_A$ as shown in Table 5. The resulting cost for this Trotterized mixer is Cost($H_M$) = 1360. Since $H \cap B$ is spanned by $k = 2^5 - |B| = 26$ computational basis states, there are $\binom{k}{2} = 325$ different pairs to add to each $T_{i+2}$. As Table 5 shows this can reduce the cost of the resulting mixer to Cost($H_M$) = 568. Of course, there is the possibility to reduce the cost even further by add more mixers for states in the kernel of $H_{M,B}$. However, this quickly becomes computationally very demanding, when all possibilities are brute-forced.

| $C$ | $T_{1+2}$ | $T_{2+3}$ | $T_{3+1}$ |
|-----|-----------|-----------|-----------|
| {}  | 12        | 8         | 16        |
| {000}, {001} | 20 | 2        | 24        |
| {000}, {101} | 24 | 20       | 28        |
| {000}, {110} | 6  | 20       | 28        |
| {000}, {111} | 28 | 24       | 8         |
| {001}, {101} | 20 | 16       | 24        |
| {001}, {110} | 28 | 24       | 8         |
| {001}, {111} | 6  | 20       | 28        |
| {101}, {110} | 24 | 20       | 28        |
| {101}, {111} | 20 | 16       | 24        |
| {110}, {111} | 20 | 2        | 24        |

Table 4: Comparison of the Cost($H_{M,B} + H_{M,C}$) for different added mixers $C$ for the case $B = \{|100\rangle, |010\rangle, |011\rangle\}$. All 10 possible pairs are shown. We see that the cost can be both reduced and increased.
6 Availability of Data and Code

All data and the python/jupyter notebook source code for reproducing the results obtained in this article are available at https://github.com/OpenQuantumComputing.

7 Conclusion and Outlook

While designing mixers with the presented framework is more or less straight forward, designing efficient mixers turns out to be a difficult tasks. An additional difficulty arises due to the need for Trotterization. Somewhat counter-intuitively, the more restricted the mixer, i.e., the smaller the subspace, the more design freedom one has to increase efficiency. More structure/symmetry of the restricted subspace seems to allow for a lower cost of the resulting mixer. For the case of “one-hot” states, we provide a deeper understanding of the requirements for Trotterizations. Compared to state of the art in literature, this leads to a considerable reduction of the cost of the mixer, as defined in Equation 35. The introduced framework reveals a rigorous mathematical analysis of the underlying structure of mixer Hamiltonians and deepens the understanding of those. We believe the framework can serve as the backbone for further development of efficient mixers.

When adding mixers, in general the kernel of $H_{M,B}$ is spanned by $k = 2^n - |B|$ computational basis states. Therefore, one can add

$$\sum_{i=2}^{k} \binom{k}{2}$$

(55)

different mixers for each non-zero entry $T_{i\rightarrow j}$ of $T$. Out of all these, one wants to find the combination leading to the lowest overall cost. Clearly, brute-forcing the optimization is computationally not tractable, even for moderate sized number of qubits $n$ when $|B| \ll 2^n$. Further research should aim to carefully analyze of the structure of the basis states in $B$ in order to develop efficient (heuristic) algorithms to find low-cost mixers through adding mixers in the kernel of $H_{M,B}$.

References

[1] Diogo Cruz et al. “Efficient quantum algorithms for GHZ and W states, and implementation on the IBM quantum computer”. In: Advanced Quantum Technologies 2.5-6 (2019), p. 1900015.

[2] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. “A quantum approximate optimization algorithm”. In: arXiv preprint arXiv:1411.4028 (2014).

[3] Franz G Fuchs et al. “Efficient Encoding of the Weighted MAX k-CUT on a Quantum Computer Using QAOA”. In: SN Computer Science 2.2 (2021), pp. 1–14.

[4] Pranav Gokhale et al. “Minimizing state preparations in variational quantum eigensolver by partitioning into commuting families”. In: arXiv preprint arXiv:1907.13623 (2019).

[5] Kaiwen Gui et al. “Term grouping and travelling salesperson for digital quantum simulation”. In: arXiv preprint arXiv:2001.05983 (2020).

[6] Stuart Hadfield et al. “From the quantum approximate optimization algorithm to a quantum alternating operator ansatz”. In: Algorithms 12.2 (2019), p. 34.

[7] Stuart Hadfield et al. “Quantum approximate optimization with hard and soft constraints”. In: Proceedings of the Second International Workshop on Post Moores Era Supercomputing. 2017, pp. 15–21.

[8] Itay Hen and Marcelo S Sarandy. “Driver Hamiltonians for constrained optimization in quantum annealing”. In: Physical Review A 93.6 (2016), p. 062312.

[9] Itay Hen and Federico M Spedalieri. “Quantum annealing for constrained optimization”. In: Physical Review Applied 5.3 (2016), p. 034007.
Elliott Lieb, Theodore Schultz, and Daniel Mattis. “Two soluble models of an antiferromagnetic chain”. In: *Annals of Physics* 16.3 (1961), pp. 407–466.

Andrew Lucas. “Ising formulations of many NP problems”. In: *Frontiers in Physics* 2 (2014), p. 5.

Pontus Vikstål et al. “Applying the Quantum Approximate Optimization Algorithm to the Tail-Assignment Problem”. In: *Phys. Rev. Applied* 14 (3 Sept. 2020), p. 034009.

Zhihui Wang et al. “XY mixers: Analytical and numerical results for the quantum alternating operator ansatz”. In: *Physical Review A* 101.1 (Jan. 2020).
Table 5: Comparison of the Cost($H_{M,B} + H_{M,C}$) for different added mixers $C$ for the case $B = \{\{10010\}, \{01110\}, \{10011\}, \{11101\}, \{00110\}, \{01010\}\}$. There are 325 possible pairs in total. We see that the cost can be both reduced and increased.