Q-FW: A Hybrid Classical-Quantum Frank-Wolfe for Quadratic Binary Optimization Supplementary Material

Alp Yurtsever[1], Tolga Birdal[2], and Vladislav Golyanik[3]

[1] Umeå University, Sweden
[2] Imperial College London, UK
[3] MPI for Informatics, SIC, Germany

Abstract. This document supplements our paper *Q-FW: A Hybrid Classical-Quantum Frank-Wolfe for Quadratic Binary Optimization*. In particular, we first provide the detailed formulations of permutation constraints used in the main paper as well as a way to factor in the inequality constraints. We then provide our convergence analysis by relating Q-FWAL to the existing literature. We also describe the early stopping heuristics and step-size choices which saved us some D-Wave time. Finally, we touch upon the gauge freedom inherent in the permutation synchronization and provide additional ablation studies along with the details of our synthetic experiments.

1 Theoretical Aspects & Discussions

1.1 Permutation-ness as a Linear Constraint

The formulation of permutation-ness into linear constraints appeared both in QGM [14] and in QuantumSync [3]. We include a brief description here for completeness. A *permutation matrix* is defined as a sparse, square binary matrix, where each column or row contains only a single non-zero entry:

$$\mathcal{P}_n := \{\mathbf{P} \in \{0,1\}^{n\times n} : \mathbf{P}\mathbf{1}_n = \mathbf{1}_n \,,\, \mathbf{1}_n^\top \mathbf{P} = \mathbf{1}_n^\top\}. \tag{1}$$

where $\mathbf{1}_n$ denotes a n-dimensional ones vector. Every $\mathbf{P} \in \mathcal{P}_n$ is a *total* permutation matrix and $P_{ij} = 1$ implies that point i is mapped to element j. Note, $\mathbf{P}^\top = \mathbf{P}^{-1}$.

During optimization, permutation-ness could be imposed on a binary vector/matrix by introducing linear constraints $\mathbf{A}\mathbf{x} = \mathbf{b}$: rows and columns sum to one as in Eq (1). Given $\mathbf{x}_i = \mathrm{vec}(\mathbf{X}_i)$, this amounts to having $\mathbf{b}_i = \mathbf{1}$ and

$$\mathbf{A}_i = \begin{bmatrix} \mathbf{I} \otimes \mathbf{1}^\top \\ \mathbf{1}^\top \otimes \mathbf{I} \end{bmatrix}. \tag{2}$$

Put simply, the matrix $\mathbf{A}_i$ is assembled as follows: in row j with $1 \leq j \leq n$, the ones are placed in columns $(j-1)\cdot n + 1$ to $(j)\cdot n$. In a row j with $j > n$, ones will be placed at $(j-n)+p\cdot n$ for $p\in\{0,...,n-1\}$. To enforce the permutation-ness of all the individual $\mathbf{x}_i$ that make up $\mathbf{x}\in\mathbb{R}^{n^2\times m}$, we construct a $n^2\times 2n$ block-diagonal matrix $\mathbf{A}=\operatorname{diag}(\mathbf{A}_1,\mathbf{A}_2,\dots,\mathbf{A}_m)$.

$$\mathbf{A}_i=\begin{bmatrix}1\,1\,0\,0\\0\,0\,1\,1\\1\,0\,1\,0\\0\,1\,0\,1\end{bmatrix},\quad \mathbf{b}_i=\mathbf{1}=\begin{bmatrix}1\\1\\1\\1\end{bmatrix}. \tag{3}$$

1.2 On Permutation Synchronization & Gauge Freedom

A close look to the presented permutation synchronization problem reveals that it is **non-convex** in $(\mathbf{P}_i,\mathbf{P}_j)$, but **convex when** one odd, *e.g.* $\mathbf{P}_i$, is fixed during optimization. In fact, if $\mathbf{P}_i$ is considered to be fixed, this problem resembles a matrix averaging under the metric the Frobenius norm.

The formulation in §5.2 is subject to a freedom in the choice of the reference or the *gauge* [1,7,3]. In other words, the solution set can be transformed arbitrarily by a common $\mathbf{P}_g$, still satisfying the consistency constraint:

$$E(\{\mathbf{X}_i\mathbf{P}_g\})=\sum_{(i,j)\in\mathcal{E}}\|\mathbf{P}_{ij}-(\mathbf{X}_i\mathbf{P}_g)(\mathbf{X}_j\mathbf{P}_g)^\top\|_{\mathrm{F}}^2 \tag{4}$$

$$=\sum_{(i,j)\in\mathcal{E}}\|\mathbf{P}_{ij}-\mathbf{X}_i\mathbf{P}_g\mathbf{P}_g^\top\mathbf{X}_j^\top\|_{\mathrm{F}}^2 \tag{5}$$

$$=\sum_{(i,j)\in\mathcal{E}}\|\mathbf{P}_{ij}-\mathbf{X}_i\mathbf{X}_j^\top\|_{\mathrm{F}}^2 \tag{6}$$

$$=E(\{\mathbf{X}_i\}). \tag{7}$$

The last equality follows from the orthogonality of permutation matrices. In practice, a gauge can be fixed by setting one of the vertex labels to identity: $\mathbf{X}_1=\mathbf{I}$. However, for convenience, we do not explicitly account for gauge freedom. We transform the first node to identity, only after obtaining the full solution.

1.3 Extended Notation

Q-FW involves a lifting procedure that maps a QBO problem with n variables and m equality constraints into a copositive program with $(n+1)^2$ variables and $2m+n+1$ equality constraints. This dimensionality expansion complicates the notation. For the ease of presentation, we introduce a compact notation in (§3) of the main text. Here, we revisit this notation for clarity.

First, we define the primal and dual dimensions $p=n+1$ and $d=2m+n+1$. Then, our primal variable is a $p\times p$ completely positive matrix $\mathbf{W}\in\Delta^p$, and our dual variable $\mathbf{y}\in\mathbb{R}^d$. $\mathbf{W}$ relates to $\mathbf{x}\in\mathbb{R}^n$ and $\mathbf{X}\in\mathbb{R}^{n\times n}$ by

$$\mathbf{W}=\begin{bmatrix}W_{11} & \mathbf{x}^\top\\ \mathbf{x} & \mathbf{X}\end{bmatrix},\ \text{with the constraint } W_{11}=1. \tag{8}$$

Then, we define a linear map $\mathcal{A} : \mathbb{R}^{p \times p} \to \mathbb{R}^d$ and $\mathbf{v} \in \mathbb{R}^d$ to simplify the writing of the constraints as $\mathcal{A}\mathbf{W} = \mathbf{v}$. Explicitly, $\mathcal{A}$ and $\mathbf{v}$ are defined by

$$\underbrace{\begin{bmatrix} W_{11} \\ X_{11} - x_1 \\ \vdots \\ X_{nn} - x_n \\ \mathbf{a}_1^\top \mathbf{x} \\ \vdots \\ \mathbf{a}_m^\top \mathbf{x} \\ \mathrm{Tr}(\mathbf{A}_1 \mathbf{X}) \\ \vdots \\ \mathrm{Tr}(\mathbf{A}_m \mathbf{X}) \end{bmatrix}}_{\mathcal{A}\mathbf{W}} = \underbrace{\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ b_1 \\ \vdots \\ b_m \\ b_1^2 \\ \vdots \\ b_m^2 \end{bmatrix}}_{\mathbf{v}} \tag{9}$$

1.4 Convergence Analysis

Our Q-FWAL algorithm is a special instance of Frank Wolfe with Augmented Lagrangian (FWAL) methods where Q-FWAL uses the QUBO-specific D-Wave solver and a tight copositive relaxation. With the observation that copositive relaxation does not have an effect on the convergence of FWAL and the assumption that the solver is *exact*[4], it is possible to consult the FW literature for a convergence analysis. We now present the proof of convergence rate of FWAL (see Proposition 1 in the main text) for completeness. The original proof appears in [15]. Our presentation closely follows the exposition in [16, Section SM1.6].

First, we exploit smoothness of L_{β_t} in the primal argument:

$$\begin{aligned} L_{\beta_t}(\mathbf{W}_{t+1}, \mathbf{y}_t) &\le L_{\beta_t}(\mathbf{W}_t, \mathbf{y}_t) + \mathrm{Tr}(\mathbf{G}_t(\mathbf{W}_{t+1} - \mathbf{W}_t)) + \frac{1}{2}\beta_t \|\mathcal{A}(\mathbf{W}_{t+1} - \mathbf{W}_t)\|_F^2 \\ &= L_{\beta_t}(\mathbf{W}_t, \mathbf{y}_t) + \eta_t \mathrm{Tr}(\mathbf{G}_t(\mathbf{H}_t - \mathbf{W}_t)) + \frac{1}{2}\beta_t \eta_t^2 \|\mathcal{A}(\mathbf{H}_t - \mathbf{W}_t)\|_F^2 \\ &\le L_{\beta_t}(\mathbf{W}_t, \mathbf{y}_t) + \eta_t \mathrm{Tr}(\mathbf{G}_t(\mathbf{H}_t - \mathbf{W}_t)) + \frac{1}{2}\beta_t \eta_t^2 \|\mathcal{A}\|^2 p^2 \\ &\le L_{\beta_t}(\mathbf{W}_t, \mathbf{y}_t) + \eta_t \mathrm{Tr}(\mathbf{G}_t(\mathbf{W}_\star - \mathbf{W}_t)) + \frac{1}{2}\beta_t \eta_t^2 \|\mathcal{A}\|^2 p^2. \end{aligned} \tag{10}$$

The second line follows by definition of $\mathbf{W}_{t+1}$, the third line holds because the Frobenius-norm diameter of Δ^p is p, and the last line depends on the fact that $\mathbf{H}_t$ minimizes $\mathrm{Tr}(\mathbf{G}_t \,\cdot\,)$.

[4] Although we do know currently that this is not true, quantum revolution might enable computers, which largely satisfy this assumption in the future.

Next, we use the definition of $\mathbf{G}_t$ to bound

$$
\begin{aligned}
\mathrm{Tr}(\mathbf{G}_t(\mathbf{W}_\star - \mathbf{W}_t)) &= \mathrm{Tr}\Big(\big(\mathbf{C} + \mathcal{A}^\top \mathbf{y}_t + \beta_t \mathcal{A}^\top(\mathcal{A}\mathbf{W}_t - \mathbf{v})\big)\big(\mathbf{W}_\star - \mathbf{W}_t\big)\Big) \\
&= \mathrm{Tr}(\mathbf{C}(\mathbf{W}_\star - \mathbf{W}_t)) + \big(\mathbf{y}_t + \beta_t(\mathcal{A}\mathbf{W}_t - \mathbf{v})\big)^\top \big(\mathcal{A}\mathbf{W}_\star - \mathcal{A}\mathbf{W}_t\big) \\
&= \mathrm{Tr}(\mathbf{C}(\mathbf{W}_\star - \mathbf{W}_t)) + \big(\mathbf{y}_t + \beta_t(\mathcal{A}\mathbf{W}_t - \mathbf{v})\big)^\top \big(\mathbf{v} - \mathcal{A}\mathbf{W}_t\big) \\
&= \mathrm{Tr}(\mathbf{C}\mathbf{W}_\star) - L_{\beta_t}(\mathbf{W}_t, \mathbf{y}_t) - \frac{\beta_t}{2}\|\mathcal{A}\mathbf{W}_t - \mathbf{v}\|^2
\end{aligned}
\tag{11}
$$

where we used the fact that $\mathcal{A}\mathbf{W}_\star = \mathbf{v}$.

We combine (10) with (11) and subtract $\mathrm{Tr}(\mathbf{C}\mathbf{W}_\star)$ to get

$$
\begin{aligned}
L_{\beta_t}(\mathbf{W}_{t+1}, \mathbf{y}_t) - \mathrm{Tr}(\mathbf{C}\mathbf{W}_\star) &\le (1-\eta_t)\Big(L_{\beta_t}(\mathbf{W}_t, \mathbf{y}_t) - \mathrm{Tr}(\mathbf{C}\mathbf{W}_\star)\Big) \\
&\quad - \frac{1}{2}\beta_t\eta_t\|\mathcal{A}\mathbf{W}_t - \mathbf{v}\|^2 + \frac{1}{2}\beta_t\eta_t^2\|\mathcal{A}\|^2 p^2.
\end{aligned}
\tag{12}
$$

Now, we update the penalty parameter on the right hand side,

$$
\begin{aligned}
L_{\beta_t}(\mathbf{W}_{t+1}, \mathbf{y}_t) - \mathrm{Tr}(\mathbf{C}\mathbf{W}_\star) &\le (1-\eta_t)\Big(L_{\beta_{t-1}}(\mathbf{W}_t, \mathbf{y}_t) - \mathrm{Tr}(\mathbf{C}\mathbf{W}_\star)\Big) \\
&\quad + \frac{1}{2}(1-\eta_t)(\beta_t - \beta_{t-1})\|\mathcal{A}\mathbf{W}_t - \mathbf{v}\|^2 \\
&\quad - \frac{1}{2}\beta_t\eta_t\|\mathcal{A}\mathbf{W}_t - \mathbf{v}\|^2 + \frac{1}{2}\beta_t\eta_t^2\|\mathcal{A}\|^2 p^2.
\end{aligned}
\tag{13}
$$

By design, our parameter choices for η_t and β_t ensures that

$$
(1-\eta_t)(\beta_t - \beta_{t-1}) \le \beta_t\eta_t. \tag{14}
$$

Therefore, we can simplify (13) to

$$
\begin{aligned}
L_{\beta_t}(\mathbf{W}_{t+1}, \mathbf{y}_t) - \mathrm{Tr}(\mathbf{C}\mathbf{W}_\star) &\le (1-\eta_t)\Big(L_{\beta_{t-1}}(\mathbf{W}_t, \mathbf{y}_t) - \mathrm{Tr}(\mathbf{C}\mathbf{W}_\star)\Big) \\
&\quad + \frac{1}{2}\beta_t\eta_t^2\|\mathcal{A}\|^2 p^2.
\end{aligned}
\tag{15}
$$

Then, we change the dual variable on the left hand side of the inequality to obtain a recursion:

$$
\begin{aligned}
&L_{\beta_t}(\mathbf{W}_{t+1}, \mathbf{y}_{t+1}) - \mathrm{Tr}(\mathbf{C}\mathbf{W}_\star) \\
&\quad = L_{\beta_t}(\mathbf{W}_{t+1}, \mathbf{y}_t) - \mathrm{Tr}(\mathbf{C}\mathbf{W}_\star) + (\mathbf{y}_{t+1} - \mathbf{y}_t)^\top(\mathcal{A}\mathbf{W}_{t+1} - \mathbf{v}) \\
&\quad = L_{\beta_t}(\mathbf{W}_{t+1}, \mathbf{y}_t) - \mathrm{Tr}(\mathbf{C}\mathbf{W}_\star) + \gamma_t\|\mathcal{A}\mathbf{W}_{t+1} - \mathbf{v}\|^2 \\
&\quad \le L_{\beta_t}(\mathbf{W}_{t+1}, \mathbf{y}_t) - \mathrm{Tr}(\mathbf{C}\mathbf{W}_\star) + \beta_t\eta_t^2\|\mathcal{A}\|^2 p^2,
\end{aligned}
\tag{16}
$$

where the last line is ensured by the assumptions on the choice of γ_t. We combine (15) and (16),

$$
\begin{aligned}
L_{\beta_t}(\mathbf{W}_{t+1}, \mathbf{y}_{t+1}) - \mathrm{Tr}(\mathbf{C}\mathbf{W}_\star) &\le (1-\eta_t)\Big(L_{\beta_{t-1}}(\mathbf{W}_t, \mathbf{y}_t) - \mathrm{Tr}(\mathbf{C}\mathbf{W}_\star)\Big) \\
&\quad + \frac{3}{2}\beta_t\eta_t^2\|\mathcal{A}\|^2 p^2.
\end{aligned}
\tag{17}
$$

Using this recursion for iterations $1 \ldots, t$, we obtain

$$L_{\beta_t}(\mathbf{W}_{t+1}, \mathbf{y}_{t+1}) - \mathrm{Tr}(\mathbf{C}\mathbf{W}_\star) \leq (1-\eta_1)\Big(L_{\beta_0}(\mathbf{W}_1, \mathbf{y}_1) - \mathrm{Tr}(\mathbf{C}\mathbf{W}_\star)\Big) + \frac{3}{2}\|\mathcal{A}\|^2 p^2 \sum_{i=1}^{t} \beta_i \eta_i^2 \prod_{j=i+1}^{t} (1-\eta_j). \tag{18}$$

The first term on the right side is 0 since $\eta_1 = 1$. We focus on the second term:

$$\begin{aligned} \sum_{i=1}^{t} \beta_i \eta_i^2 \prod_{j=i+1}^{t} (1-\eta_j) &= 4\beta_0 \sum_{i=1}^{t} \frac{1}{(i+1)^{3/2}} \prod_{j=i+1}^{t} \frac{j-1}{j+1} \\ &= 4\beta_0 \sum_{i=1}^{t} \frac{1}{(i+1)^{3/2}} \frac{i(i+1)}{t(t+1)} \leq \frac{4\beta_0}{t(t+1)} \sum_{i=1}^{t} i^{1/2} \leq \frac{4\beta_0}{\sqrt{t+1}}. \end{aligned} \tag{19}$$

Hence, we conclude that

$$L_{\beta_t}(\mathbf{W}_{t+1}, \mathbf{y}_{t+1}) - \mathrm{Tr}(\mathbf{C}\mathbf{W}_\star) \leq \frac{6\beta_0 \|\mathcal{A}\|^2 p^2}{\sqrt{t+1}}. \tag{20}$$

The bound on the objective residual follows immediately from (20), since

$$\begin{aligned} L_{\beta_t}(\mathbf{W}_{t+1}, \mathbf{y}_{t+1}) &= \mathrm{Tr}(\mathbf{C}\mathbf{W}_{t+1}) + \mathbf{y}_{t+1}^\top(\mathcal{A}\mathbf{W}_{t+1} - \mathbf{v}) + \frac{\beta_t}{2}\|\mathcal{A}\mathbf{W}_{t+1} + \mathbf{v}\|^2 \\ &= \mathrm{Tr}(\mathbf{C}\mathbf{W}_{t+1}) - \frac{1}{2\beta_t}\|\mathbf{y}_{t+1}\|^2 + \frac{\beta_t}{2}\|\mathcal{A}\mathbf{W}_{t+1} + \mathbf{v} - \beta_t^{-1}\mathbf{y}\|^2 \\ &\geq \mathrm{Tr}(\mathbf{C}\mathbf{W}_{t+1}) - \frac{D^2}{2\beta_t}, \end{aligned} \tag{21}$$

where the last line depends on the boundedness assumption on $\mathbf{y}_t$. We combine (20) and (21) and get

$$\mathrm{Tr}(\mathbf{C}\mathbf{W}_{t+1}) - \mathrm{Tr}(\mathbf{C}\mathbf{W}_\star) \leq \frac{6\beta_0 \|\mathcal{A}\|^2 p^2}{\sqrt{t+1}} + \frac{D^2}{2\beta_0\sqrt{t+1}}. \tag{22}$$

It remains to prove the bound on infeasibility. We revisit (22), invoke Cauchy-Schwarz inequality and the boundedness assumption on $\mathbf{y}$ to obtain

$$\begin{aligned} &L_{\beta_t}(\mathbf{W}_{t+1}, \mathbf{y}_{t+1}) - \mathrm{Tr}(\mathbf{C}\mathbf{W}_\star) \\ &= \mathrm{Tr}(\mathbf{C}\mathbf{W}_{t+1}) - \mathrm{Tr}(\mathbf{C}\mathbf{W}_\star) + \mathbf{y}_{t+1}^\top(\mathcal{A}\mathbf{W}_{t+1} - \mathbf{v}) + \frac{\beta_t}{2}\|\mathcal{A}\mathbf{W}_{t+1} + \mathbf{v}\|^2 \\ &\leq \frac{6\beta_0 \|\mathcal{A}\|^2 p^2}{\sqrt{t+1}}. \end{aligned} \tag{23}$$

Based on the strong duality assumption, we use the Lagrangian saddle point theory [5, Section 5.4],

$$\underbrace{\mathrm{Tr}(\mathbf{C}\mathbf{W}_\star)}_{L_0(\mathbf{W}_\star, \mathbf{y}_\star)} \leq \underbrace{\mathrm{Tr}(\mathbf{C}\mathbf{W}_{t+1}) + \mathbf{y}_\star^\top(\mathcal{A}\mathbf{W}_{t+1} - \mathbf{v})}_{L_0(\mathbf{W}_{t+1}, \mathbf{y}_\star)}. \tag{24}$$

We combine (23) and (24):

$$(\mathbf{y}_{t+1} - \mathbf{y}_\star)^\top (\mathcal{A}\mathbf{W}_{t+1} - \mathbf{v}) + \frac{\beta_t}{2}\|\mathcal{A}\mathbf{W}_{t+1} + \mathbf{v}\|^2 \leq \frac{6\beta_0\|\mathcal{A}\|^2 p^2}{\sqrt{t+1}}. \tag{25}$$

We use Cauchy-Schwarz and the boundedness assumption on $\mathbf{y}$ to obtain a second-order inequality of $\|\mathcal{A}\mathbf{W}_{t+1} - \mathbf{v}\|$:

$$-2D\|\mathcal{A}\mathbf{W}_{t+1} - \mathbf{v}\| + \frac{\beta_t}{2}\|\mathcal{A}\mathbf{W}_{t+1} + \mathbf{v}\|^2 \leq \frac{6\beta_0\|\mathcal{A}\|^2 p^2}{\sqrt{t+1}}. \tag{26}$$

By solving this inequality for $\|\mathcal{A}\mathbf{W}_{t+1} - \mathbf{v}\| \geq 0$, we get

$$\|\mathcal{A}\mathbf{W}_{t+1} - \mathbf{v}\| \leq \frac{1}{\beta_t}\left(4D + 2\sqrt{3}\beta_0 p\|\mathcal{A}\|\right). \tag{27}$$

1.5 On the Dual Step-size of FWAL

Theoretical analysis of FWAL depends on the assumption that the dual step-size $\gamma_t \geq 0$ satisfies

$$\gamma_t\|\mathbf{g}_t\|^2 \leq \beta_t\eta_t^2 p^2\|\mathcal{A}\|^2 \tag{28}$$

and the bounded travel condition $\|\mathbf{y}_{t+1}\| \leq D$. Note, the largest step-size that satisfies this condition can be computed analytically. First, we take the square,

$$\|\mathbf{y}_{t+1}\|^2 = \|\mathbf{y}_t + \gamma_t\mathbf{g}_t\|^2 = \|\mathbf{y}_t\|^2 + 2\gamma_t\mathbf{y}_t^\top\mathbf{g}_t + \gamma_t^2\|\mathbf{g}_t\|^2 \leq D^2. \tag{29}$$

Then, by solving this inequality for $\gamma_t \geq 0$, we obtain

$$\gamma_t \leq \frac{-\mathbf{y}_t^\top\mathbf{g}_t + \sqrt{(\mathbf{y}_t^\top\mathbf{g}_t)^2 + (D^2 - \|\mathbf{y}_t\|^2)\|\mathbf{g}_t\|^2}}{\|\mathbf{g}_t\|^2}. \tag{30}$$

Combining (28) and (30), we can choose

$$\gamma_t \leq \min\left\{\frac{\beta_t\eta_t^2 p^2\|\mathcal{A}\|^2}{\|\mathbf{g}_t\|^2}, \frac{-\mathbf{y}_t^\top\mathbf{g}_t + \sqrt{(\mathbf{y}_t^\top\mathbf{g}_t)^2 + (D^2 - \|\mathbf{y}_t\|^2)\|\mathbf{g}_t\|^2}}{\|\mathbf{g}_t\|^2}\right\} \tag{31}$$

and $\gamma_t = 0$ if $\|\mathbf{g}_t\| = 0$. Prior work [15,16] invoke also the fixed threshold $\gamma_t \leq \beta_0$ to avoid very large steps when $\|\mathbf{g}_t\|$ is small:

$$\gamma_t \leq \min\left\{\beta_0, \frac{\beta_t\eta_t^2 p^2\|\mathcal{A}\|^2}{\|\mathbf{g}_t\|^2}, \frac{-\mathbf{y}_t^\top\mathbf{g}_t + \sqrt{(\mathbf{y}_t^\top\mathbf{g}_t)^2 + (D^2 - \|\mathbf{y}_t\|^2)\|\mathbf{g}_t\|^2}}{\|\mathbf{g}_t\|^2}\right\}. \tag{32}$$

Note, $\gamma_t = 0$ always satisfies this condition hence is a valid choice. In fact, FWQP is a special case of FWAL with $\mathbf{y}_0 = \mathbf{0}$ and $\gamma_t = 0$.

In numerical experiments, we use a constant step-size $\gamma_t = \beta_0$. This choice may fail the conditions in (32) but works well in practice.

1.6 Inequality Constraints

The problems addressed in our paper are concerned with equality constraints. However, many problems such as resolving *partial* permutations might require us to naturally handle inequalities. In this section we present one possible way to accommodate affine inequality constraints in Q-FW and leave it as a future study to experiment on tasks with such constraints. Our particular solution requires the evaluation of D-Wave Quantum Computer only as many times as in the case of equality constraints.

Without loss of generality, we assume that the inequality constraints are given in the form of

$$\mathbf{e}_i^\top \mathbf{x} \leq f_i, \quad i = 1, 2, \ldots, q. \tag{33}$$

Since $\mathbf{x}$ is binary valued, we can derive trivial lower and upper bounds

$$\begin{gathered} -\alpha_i \leq \mathbf{e}_i^\top \mathbf{x} \leq \beta_i, \quad \text{where} \\ \alpha_i = -\sum_{j=1}^{n} \min\{(\mathbf{e}_i)_j, 0\}, \quad \text{and} \quad \beta_i = \sum_{j=1}^{n} \max\{(\mathbf{e}_i)_j, 0\}. \end{gathered} \tag{34}$$

In other words, α_i is the sum of absolute values of the negative coefficients of $\mathbf{e}_i$, and β_i is the sum of its positive coefficients. By definition, α_i and β_i are nonnegative. We assume $f_i < \beta_i$, because otherwise the constraint is redundant and we can remove it. We also assume that $-\alpha_i \leq f_i$. Otherwise, the feasible set is empty and there are no solutions.

We combine (33) and (34), add α_i to both sides:

$$0 \leq \mathbf{e}_i^\top \mathbf{x} + \alpha_i \leq f_i + \alpha_i, \quad i = 1, 2, \ldots, q. \tag{35}$$

Since all sides are nonnegative, we can now take the squares and get

$$0 \leq (\mathbf{e}_i^\top \mathbf{x})^2 + \alpha_i^2 + 2\alpha_i(\mathbf{e}_i^\top \mathbf{x}) \leq (f_i + \alpha_i)^2, \quad i = 1, 2, \ldots, q. \tag{36}$$

Then, we replace $(\mathbf{e}_i^\top \mathbf{x})^2 = \mathrm{Tr}(\mathbf{x}^\top \mathbf{e}_i \mathbf{e}_i^\top \mathbf{x}) = \mathrm{Tr}(\mathbf{e}_i \mathbf{e}_i^\top \mathbf{x}\mathbf{x}^\top)$ by $\mathrm{Tr}(\mathbf{E}_i \mathbf{X})$ where $\mathbf{E}_i = \mathbf{e}_i \mathbf{e}_i^\top$. We subtract α_i^2 and get

$$-\alpha_i^2 \leq \mathrm{Tr}(\mathbf{E}_i \mathbf{X}) + 2\alpha_i(\mathbf{e}_i^\top \mathbf{x}) \leq f_i^2 + 2\alpha_i f_i, \quad i = 1, 2, \ldots, q. \tag{37}$$

By combining (35) and (37), we reformulate q inequality constraints of the original QBO problem as $2q$ inequality (box) constraints in our CP relaxation:

$$\begin{aligned} -\alpha_i &\leq \mathbf{e}_i^\top \mathbf{x} \leq f_i, && i = 1, 2, \ldots, q \\ -\alpha_i^2 &\leq \mathrm{Tr}(\mathbf{E}_i \mathbf{X}) + 2\alpha_i(\mathbf{e}_i^\top \mathbf{x}) \leq f_i^2 + 2\alpha_i f_i, && i = 1, 2, \ldots, q \end{aligned} \tag{38}$$

Finally, we introduce a linear map $\mathcal{E} : \mathbb{R}^{p\times p} \to \mathbb{R}^{2q}$ and two vectors $\boldsymbol{l}, \mathbf{u} \in \mathbb{R}^{2q}$ to simplify the notation to $\boldsymbol{l} \leq \mathcal{E}(\mathbf{W}) \leq \mathbf{u}$, or explicitly,

$$\underbrace{\begin{bmatrix} -\alpha_1 \\ \vdots \\ -\alpha_q \\ -\alpha_1^2 \\ \vdots \\ -\alpha_q^2 \end{bmatrix}}_{\boldsymbol{l}} \leq \underbrace{\begin{bmatrix} \mathbf{e}_1^\top \mathbf{x} \\ \vdots \\ \mathbf{e}_q^\top \mathbf{x} \\ \mathrm{Tr}(\mathbf{E}_1\mathbf{X}) + 2\alpha_1(\mathbf{e}_1^\top \mathbf{x}) \\ \vdots \\ \mathrm{Tr}(\mathbf{E}_q\mathbf{X}) + 2\alpha_q(\mathbf{e}_q^\top \mathbf{x}) \end{bmatrix}}_{\mathcal{E}(\mathbf{W})} \leq \underbrace{\begin{bmatrix} f_1 \\ \vdots \\ f_q \\ f_1^2 + 2\alpha_1 f_1 \\ \vdots \\ f_q^2 + 2\alpha_q f_q \end{bmatrix}}_{\mathbf{u}} \tag{39}$$

where the inequalities are entrywise.

Next, we present FWAL steps for inequality constraints. This extension is detailed in [16, Section D.4], we present it here for completeness. We use the following augmented Lagrangian formulation to derive FWAL steps:

$$\begin{aligned} L_\beta(\mathbf{W}; \mathbf{y}; \mathbf{y}') = \mathrm{Tr}(\mathbf{C}\mathbf{W}) + \mathbf{y}^\top(\mathcal{A}\mathbf{W} - \mathbf{v}) + \frac{\beta}{2}\|\mathcal{A}\mathbf{W} - \mathbf{v}\|^2 \\ + \min_{\boldsymbol{l} \leq \boldsymbol{\omega} \leq \mathbf{u}} \left\{ \mathbf{y}'^\top(\mathcal{E}\mathbf{W} - \boldsymbol{\omega}) + \frac{\beta}{2}\|\mathcal{E}\mathbf{W} - \boldsymbol{\omega}\|^2 \right\}. \end{aligned} \tag{40}$$

Then, the partial derivative of L_{β_t} with respect to the first variable is

$$\begin{aligned} \mathbf{G}_t = \mathbf{C} + \mathcal{A}^\top \mathbf{y}_t + \beta_t \mathcal{A}^\top(\mathcal{A}\mathbf{W}_t - \mathbf{v}) + \mathcal{E}^\top \mathbf{y}'_t + \beta_t \mathcal{E}^\top(\mathcal{E}\mathbf{W}_t - \boldsymbol{\omega}_t^\star), \\ \text{where } \boldsymbol{\omega}_t^\star = \arg\min_{\boldsymbol{l} \leq \boldsymbol{\omega} \leq \mathbf{u}} \left\{ {\mathbf{y}'_t}^\top(\mathcal{E}\mathbf{W}_t - \boldsymbol{\omega}) + \frac{\beta_t}{2}\|\mathcal{E}\mathbf{W}_t - \boldsymbol{\omega}\|^2 \right\}. \end{aligned} \tag{41}$$

The $\boldsymbol{\omega}_t^\star$ subproblem amounts to a projection, which in turn is a clipping (thresholding) operator:

$$\begin{aligned} \boldsymbol{\omega}_t^\star &= \arg\min_{\boldsymbol{l} \leq \boldsymbol{\omega} \leq \mathbf{u}} \left\{ {\mathbf{y}'_t}^\top(\mathcal{E}\mathbf{W}_t - \boldsymbol{\omega}) + \frac{\beta_t}{2}\|\mathcal{E}\mathbf{W}_t - \boldsymbol{\omega}\|^2 \right\} \\ &= \arg\min_{\boldsymbol{l} \leq \boldsymbol{\omega} \leq \mathbf{u}} \left\{ \|\mathcal{E}\mathbf{W}_t - \boldsymbol{\omega} + \beta_t^{-1}\mathbf{y}'_t\|^2 \right\} \\ &= \mathrm{proj}_{[\boldsymbol{l},\mathbf{u}]}(\mathcal{E}\mathbf{W}_t + \beta_t^{-1}\mathbf{y}'_t) := \mathrm{clip}(\mathcal{E}\mathbf{W}_t + \beta_t^{-1}\mathbf{y}'_t, \boldsymbol{l}, \mathbf{u}). \end{aligned} \tag{42}$$

The update rule for the dual variable $\mathbf{y}$ remains the same. Similarly, for $\mathbf{y}'$, we take a small gradient ascent step by using the partial derivative of L_β with respect to the third variable,

$$\mathbf{g}'_t = \mathcal{E}\mathbf{W}_{t+1} - \boldsymbol{\omega}_t^\star, \quad \text{and} \quad \mathbf{y}'_{t+1} = \mathbf{y}'_t + \gamma_t \mathbf{g}'_t. \tag{43}$$

1.7 Early Stopping Heuristics

As D-Wave provides a limited amount of computation, we are bound to use our resources wisely. To this end, for some of the synchronization experiments[5], we

[5] usually for QPS, we solve larger problems than QGGM

opt for (i) a faster update, (ii) an automatic termination when good quality solutions are found. We take a different approach and propose two modifications to the original Q-FWAL:

1. projecting the solution to the feasibility set at each iteration and switching the current solution with the projected, if:
$$\mathrm{Tr}(\mathbf{C}\hat{\mathbf{H}}_t) < \mathrm{Tr}(\mathbf{C}\mathbf{W}_t) \tag{44}$$
where $\hat{\mathbf{H}}_t$ is obtained by lifting the rounded, intermediate solution at time t, *i.e.* for permutations, applying Hungarian algorithm on the left singular vectors of $\mathbf{X}_t$.
2. the stopping criteria that checks the constraints are satisfied and the cost remains unchanged in consecutive iterations:
$$\|\mathcal{A}\mathbf{W}_t - \mathbf{v}\| = 0 \quad \text{and} \quad \mathrm{Tr}(\mathbf{C}\mathbf{W}_t) = \mathrm{Tr}(\mathbf{C}\mathbf{W}_{t-1}) \tag{45}$$

Note that, typical Frank Wolfe-type algorithms usually make use of the *duality gap* as a practical *stopping criterion* motivated by the fact that this quantity upper bounds the primal gap while at the same time enjoying the same asymptotic guarantees. [10,11,16]. However, we find that in practice this is still a very soft barrier, satisfied only at high number of iterations. This is the reason why we preferred the two proposed modifications above.

2 Adiabatic Quantum Computing

Adiabatic quantum computing (AQC) is only one of the two quantum computing models. AQC and gate-based quantum computing paradigms are said to be polynomially equivalent, in theory (experimental confirmations are ongoing). In the gate-based model, all computations on qubits can be represented as unitary transformations (that can potentially cover the entire Hilbert space); hence, all operations before qubit measurements are invertible. AQC model, instead, is defined in terms of Hamilton operator evolution. Note, QA can be performed both in an adiabatic and non-adiabatic manner (faster than what the adiabatic theorem dictates). Current AQC implementations such as DWave [8] implement QA, and the quantum system evolution is not guaranteed to be adiabatic. For a more comprehensive overview of the AQC foundations, see [12,9,13].

The weight matrix of a QUBO problem defines a *logical* problem [2], *i.e.,* each its binary variable is said to be a *logical* qubit in the idealised quantum hardware context. Every logical problem is abstracted from real quantum hardware and assumes arbitrary connectivity patterns between the qubits. This contrasts with the notion of *physical* qubits, *i.e.,* qubits available in hardware with their connectivity patterns. Since physical qubits are not arbitrarily connected to each other on modern AQCs, multiple of them are required to represent a single logical problem qubit [8]. Finding a mapping of a logical QUBO problem to the hardware qubit graph is known as *minor embedding*; it can be performed with such algorithms as Cai *et al.* [6]. We give an example involving logical and

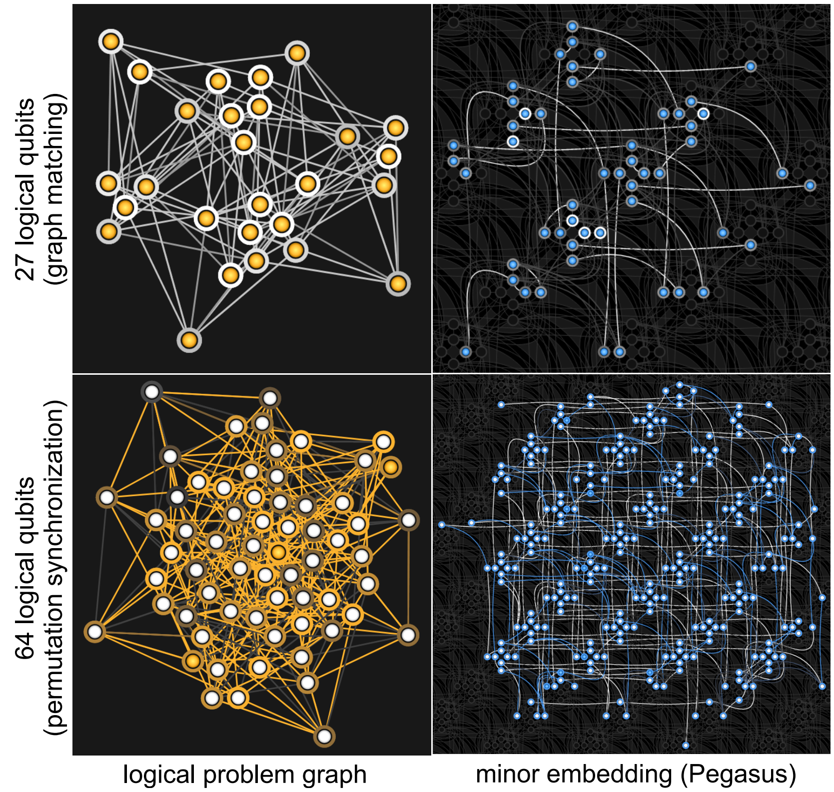

Fig. 1: Graphs of the logical problems (the left column) arising in our experiments with 27 (the top row) and 64 logical qubits (the bottom row), along with their minor embeddings on the Pegasus topology [8] obtained by Cai *et al.*'s method [6] (the right column). Each node in the logical problem graph represents a logical qubit, and each edge stands for couplings between the logical qubits. Physical qubits build chains in the minor embedding to represent a single logical qubit.

embedded graphs of two of our problems in Fig. 1. We now briefly describe quantum annealing.

A QUBO optimization is equivalent to minimizing the energy of a classical Ising Hamiltonian $\mathbf{J}$ with no bias field, where the variables $\mathbf{s}_i$ are interpreted as classical spin values. Hence, the minimum of the QUBO objective is equivalently obtained as the *ground state* of Quantum Ising Hamiltonian:

$$\mathcal{H}_P = \sum_{ij} J_{ij} \sigma_P^{(i)} \sigma_P^{(j)}, \tag{46}$$

where $\sigma_P^{(i)}$ denotes the *Pauli matrix* applied to the i^{th} qubit of an n-qubit system. In contrast to a classical bit, a qubit $|\psi\rangle$ can continuously transition between the states $|0\rangle$ and $|1\rangle$ (the equivalents of classical states 0 and 1) fulfilling the equation $|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle$, with probability amplitudes satisfying $|\alpha|^2 + |\beta|^2 = 1$. The eigenvalues of the Hamiltonian correspond to the possible system's energies. As such, the same minimization can be written as:

$$\min_{|\psi\rangle\in\mathbb{C}^{2^n}} \langle\psi|\,\mathcal{H}_P\,|\psi\rangle\,. \tag{47}$$

Adiabatic Quantum Annealing (AQA) solves Eq (47) by evolving the Hamiltonian to one where the ground state corresponds to the optimal solution:

$$\mathcal{H}(\tau) = [1-\tau]\,\mathcal{H}_I + \tau\,\mathcal{H}_P, \tag{48}$$

with $\mathcal{H}_I$ being an *initial* Hamiltonian realized as a superposition with equal probabilities of measuring $|0\rangle$ or $|1\rangle$ for every qubit. The adiabatic theorem of quantum mechanics [4] implies that if a system transits *gradually enough* (the concrete meaning of *gradually* depends on many factors), then the system will continue to stay in its ground state in the course of the entire evolution. Hence, by the end of the transition, the system will be measured in the ground state of the problem Hamiltonian, *i.e.,* the global optimiser.

A hybrid algorithm involving QA always has multiple steps that cover the preparation of a QUBO problem, minor embedding, a series of anneals, problem unembedding (from the graph of physical qubits to the logical problem graph), solution selection and solution interpretation.

2.1 Psuedocode

We are now ready to provide the pseudocode for Q-FW. In the main paper we always use the equality constraints as these are the most common in the tasks we address. However, for the sake of generality we present in Alg. 1 the generic Q-FW approach for handling inequality and equality constraints. We will make our implementation available upon publication.

3 Additional Evaluations

On synthetic data. Our synthetic data serves the purpose of being able to execute our exhaustive binary solver for obtaining a globally optimal solution to QUBO small problems ($n = 3$ and $m = 3$). Similar to [3], we visualize in Fig. 2, some noisy and noise-free examples from our random synthetic dataset. For illustration purposes we show the case of $n = 4$ and $m = 4$, although we used smaller problems in the experiments. The important cues are the correspondences denoting permutations, whose rows might be randomly swapped to inject noise.

On the evolution of sub-problems & sparsity. We now visually compare the sub-problems emerging in solving the noiseless, synthetic synchronization problem (detailed in the previous experiment and in our supplementary material),

Algorithm 1 Q-FW for Quadratic Binary Optimization.

Input: Cost matrix $\mathbf{Q} \in \mathbb{R}^{n\times n}$, Equality constraints $\{(\mathbf{a}_i, b_i)\}_{i=1}^m$, Inequality constraints $\{(\mathbf{e}_i, f_i)\}_{i=1}^q$, # of iterations T, penalty parameter $\beta_0 > 0$ (default 1)

Preparation: $p \leftarrow n+1,\ d \leftarrow 2m+n+1,\ d' \leftarrow 2q$. Form $\mathbf{C} \leftarrow \left[\begin{smallmatrix} 0 & \mathbf{0}^\top \\ \mathbf{0} & \mathbf{Q} \end{smallmatrix}\right]$. Construct $(\mathcal{A}, \mathbf{v})$ as defined in (9), and $(\mathcal{E}, \boldsymbol{l}, \mathbf{u})$ as in (39).

Initialization: $\mathbf{W} \leftarrow \mathbf{0}^{p\times p},\ \mathbf{y} \leftarrow \mathbf{0}^d,\ \mathbf{y}' \leftarrow \mathbf{0}^{d'}$

Main loop [FWAL]:

for $t = 1, \dots, T$ **do**

  $\eta \leftarrow 2/(t+1)$, and $\beta \leftarrow \beta_0\sqrt{t+1}$

  $\mathbf{g} \leftarrow \mathcal{A}\mathbf{W} - \mathbf{v}$, and $\mathbf{g}' \leftarrow \mathcal{E}\mathbf{W} - \mathrm{clip}(\mathcal{E}\mathbf{W} + \beta^{-1}\mathbf{y}', \boldsymbol{l}, \mathbf{u})$

  $\mathbf{G} \leftarrow \mathbf{C} + \mathcal{A}^\top(\mathbf{y} + \beta\mathbf{g}) + \mathcal{E}^\top(\mathbf{y}' + \beta\mathbf{g}')$

  $\mathbf{w} \leftarrow \arg\min\,\{\mathbf{w}^\top\mathbf{G}\mathbf{w} :\ \mathbf{w} \in \mathbb{Z}_2^p\}$ `// QUBO subproblem`

  $\mathbf{W} \leftarrow (1-\eta)\mathbf{W} + \eta\mathbf{w}\mathbf{w}^\top$

  $\mathbf{g} \leftarrow \mathcal{A}\mathbf{W} - \mathbf{v}$, and $\mathbf{g}' \leftarrow \mathcal{E}\mathbf{W} - \mathrm{clip}(\mathcal{E}\mathbf{W} + \beta_+^{-1}\mathbf{y}', \boldsymbol{l}, \mathbf{u})$ `//` $\beta_+ = \beta_0\sqrt{t+2}$

  $\mathbf{y} \leftarrow \mathbf{y} + \gamma\mathbf{g}$, and $\mathbf{y}' \leftarrow \mathbf{y}' + \gamma\mathbf{g}'$ `// In practice, we use` $\gamma = \beta_0$

Main loop [FWQP]:

for $t = 1, \dots, T$ **do**

  $\eta \leftarrow 2/(t+1)$, and $\beta \leftarrow \beta_0\sqrt{t+1}$

  $\mathbf{g} \leftarrow \mathcal{A}\mathbf{W} - \mathbf{v}$, and $\mathbf{g}' \leftarrow \mathcal{E}\mathbf{W} - \mathrm{clip}(\mathcal{E}\mathbf{W}, \boldsymbol{l}, \mathbf{u})$

  $\mathbf{G} \leftarrow \mathbf{C} + \beta\mathcal{A}^\top\mathbf{g} + \beta\mathcal{E}^\top\mathbf{g}'$

  $\mathbf{w} \leftarrow \arg\min\,\{\mathbf{w}^\top\mathbf{G}\mathbf{w} :\ \mathbf{w} \in \mathbb{Z}_2^p\}$ `// QUBO subproblem`

  $\mathbf{W} \leftarrow (1-\eta)\mathbf{W} + \eta\mathbf{w}\mathbf{w}^\top$

Rounding: (Option 1) Extract $\mathbf{x}$ by taking the first column of $\mathbf{W}$ and removing its first entry. (Option 2) Extract $\mathbf{X}$ by removing the first row and first column of $\mathbf{W}$. Compute $\mathbf{x}$ as the top singular vector of $\mathbf{X}$. – Project $\mathbf{x}$ onto $\mathbb{Z}_2^n$.

Output: Solution $\mathbf{W} \in \Delta^p$ for the copositive program, and $\mathbf{x} \in \mathbb{Z}_2^n$ for the QBO.

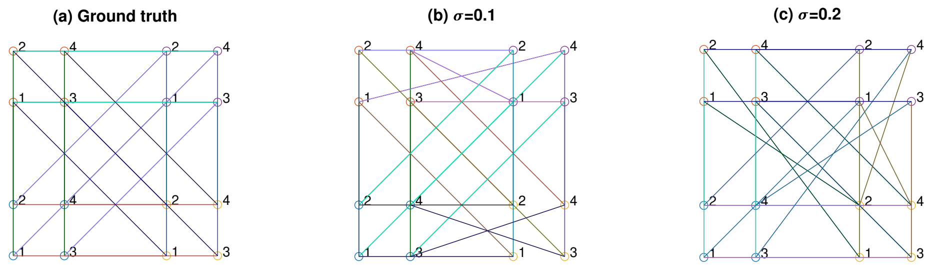

Fig. 2: Samples from our fully connected synthetic dataset for different values of swap ratio σ. In the figure, each group (indicated using differently colored points) corresponds to a *view* and each inter-group correspondence corresponds to a permutation that we optimize for. Note that the points are drawn as a grid to ease visual perception. Neither our algorithm nor the state-of-the-art methods we compare would use this information.

for our exact method and for the D-Wave implementation. As seen in Fig. 3, there is no noticeable difference between the two evolutions, confirming that D-Wave could solve the sub-QUBO-problems reliably. Moreover, over iterations the sparsity pattern of $\mathbf{W}_t$ is fixed, which means that we could compute the minor embedding[6], and re-use it throughout Q-FW. This ability of avoiding repetitive minor embeddings is a by-product of our approach and makes it a practically feasible algorithm.

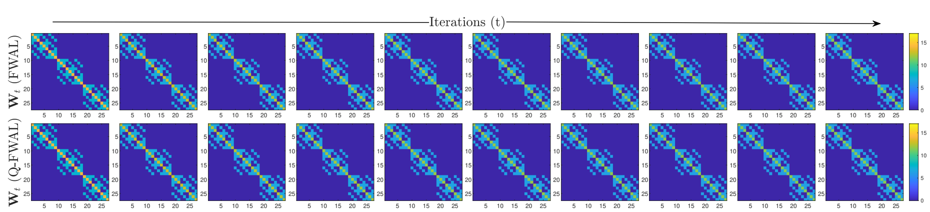

Fig. 3: Evolution of the gradient $\mathbf{W}_t$ for $0 < t < 100$ sampled in steps of 10: FWAL (top) and Q-FWAL (bottom).

References

1. Belov, V.: On geometry and symmetries in classical and quantum theories of gauge gravity. arXiv:1905.06931 (2019) 2

[6] Finding a minor embedding requires solving a combinatorial optimization problem with heuristics [6]

2. Benkner, M.S., Lähner, Z., Golyanik, V., Wunderlich, C., Theobalt, C., Moeller, M.: Q-match: Iterative shape matching via quantum annealing. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 7586–7596 (2021) 9
3. Birdal, T., Golyanik, V., Theobalt, C., Guibas, L.J.: Quantum permutation synchronization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 13122–13133 (2021) 1, 2, 11
4. Born, M., Fock, V.: Beweis des adiabatensatzes. Zeitschrift für Physik **51**(3), 165–180 (1928) 11
5. Boyd, S., Boyd, S.P., Vandenberghe, L.: Convex optimization. Cambridge university press (2004) 5
6. Cai, J., Macready, W.G., Roy, A.: A practical heuristic for finding graph minors. arXiv e-prints (2014) 9, 10, 13
7. Chatterjee, A., Govindu, V.M.: Efficient and robust large-scale rotation averaging. In: International Conference on Computer Vision (ICCV). pp. 521–528 (2013) 2
8. Dattani, N., Szalay, S., Chancellor, N.: Pegasus: The second connectivity graph for large-scale quantum annealing hardware. arXiv e-prints (2019) 9, 10
9. Farhi, E., Goldstone, J., Gutmann, S., Lapan, J., Lundgren, A., Preda, D.: A quantum adiabatic evolution algorithm applied to random instances of an np-complete problem. Science **292**(5516), 472–475 (2001) 9
10. Frandi, E., Ñanculef, R., Suykens, J.A.: A partan-accelerated frank-wolfe algorithm for large-scale svm classification. In: 2015 International Joint Conference on Neural Networks (IJCNN). pp. 1–8. IEEE (2015) 9
11. Jaggi, M.: Revisiting frank-wolfe: Projection-free sparse convex optimization. In: International Conference on Machine Learning. pp. 427–435. PMLR (2013) 9
12. Kadowaki, T., Nishimori, H.: Quantum annealing in the transverse ising model. Phys. Rev. E **58**, 5355–5363 (1998) 9
13. Technical description of the d-wave quantum processing unit. `https://docs.dwavesys.com/docs/latest/doc_qpu.html`, accessed on the 05.03.2022 9
14. Seelbach Benkner, M., Golyanik, V., Theobalt, C., Moeller, M.: Adiabatic quantum graph matching with permutation matrix constraints. In: International Conference on 3D Vision (3DV) (2020) 1
15. Yurtsever, A., Fercoq, O., Cevher, V.: A conditional-gradient-based augmented lagrangian framework. In: International Conference on Machine Learning. pp. 7272–7281. PMLR (2019) 3, 6
16. Yurtsever, A., Tropp, J.A., Fercoq, O., Udell, M., Cevher, V.: Scalable semidefinite programming. SIAM Journal on Mathematics of Data Science **3**(1), 171–200 (2021) 3, 6, 8, 9