

©2024 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

A Fast Algorithm for Computing the Deficiency Number of a Mahjong Hand

Xueqing Yan, Yongming Li, Sanjiang Li

Abstract—The tile-based multiplayer game Mahjong is widely played in Asia and has also become increasingly popular worldwide. Face-to-face or online, each player begins with a hand of 13 tiles and players draw and discard tiles in turn until they complete a winning hand. An important notion in Mahjong is the *deficiency number* (a.k.a. *shanten* number in Japanese Mahjong) of a hand, which estimates how many tile changes are necessary to complete the hand into a winning hand. The deficiency number plays an essential role in major decision-making tasks such as selecting a tile to discard. This paper proposes a fast algorithm for computing the deficiency number of a Mahjong hand. Compared with the baseline algorithm, the new algorithm is usually 100 times faster and, more importantly, respects the agent’s knowledge about available tiles. The algorithm can be used as a basic procedure in all Mahjong variants by both rule-based and machine learning-based Mahjong AI.

I. INTRODUCTION

Games have played as the test-beds of many novel artificial intelligence (AI) techniques and ideas since the very beginning of AI research. In the past decades, we have seen AI programs that can beat best human players in perfect information games including checker [10], chess [11] and Go [12], where players know everything occurred in the game before making a decision. More recently, important progress has also been made in solving the more challenging imperfect information games such as the two-player heads-up Texas hold’em poker [2], [3], DouDiZhu [5], and Mahjong [7].

In this paper, we study the imperfect information game Mahjong, which is a widely played multiplayer game. Compared with Go and poker, Mahjong is perhaps more popular and more complicated. The game is played with a set of 144 tiles based on Chinese characters and symbols and has many variants in tiles used, rules, and scoring systems [1]. In the beginning of the game, each player is given with a hand of 13 tiles and, in turn, players draw and discard tiles until they complete a winning hand.

Most early researches focus on developing AI programs for Japanese Mahjong and are often reported in Japanese. Mizukami and Tsuruoka [9] build a strong Mahjong AI based on Monte Carlo simulation and opponent models. Yoshimura et al. [14] propose a tabu search method of optimal movements without using game records. More recently, Kurita and Hoki [6] represent the game of Mahjong as multiple Markov Decision Processes and the thus constructed Mahjong AI was evaluated to be competitive with two strong AI players. Suphx, the Mahjong AI from Microsoft Research Asia [7], is by far the strongest AI for Japanese Mahjong. Based on deep reinforcement learning, Suphx has demonstrated super-human performance in the Tenhou platform (<https://tenhou.net/>).

There are also some very recent researches on Chinese Mahjong. In [13], Wang et al. propose a deep residual network-based strategy that is compared favourably with three high-level human players. Gao and Li [4] extract Mahjong features with the deep learning model and then derive their strategy by classifying the learned features. More progresses are also reported in the IJCAI 2020 Mahjong AI Competition.¹

While these achievements are remarkable, they (except [14]) often demand high volume game records of human expert players. In addition, like general deep learning algorithms, most of these Mahjong AIs have the limitation of poor interpretability. In a previous paper [8], we initiated a mathematical and knowledge-based AI study of Mahjong, aiming to explore ways of integrating traditional knowledge-based AI with machine learning. As a combinatorial game, winning hands of Mahjong have structures that can be exploited and specified as local constraints. Moreover, the agent’s belief or knowledge about the distribution of non-shown tiles can be updated as knowledge/belief revision. We expect that knowledge reasoning methods, e.g., Bayesian inference, constraint solving and optimisation, belief revision, can be exploited in the design of strong Mahjong AI. It is also expected that these knowledge-based methods can be further combined with machine learning methods proposed in [2], [3], [7].

To build a strong Mahjong AI, one basic procedure is to determine how many tile changes we need to complete the hand. This notion is called the deficiency number of the hand in [8], which is exactly the *shanten* number (or simply *shanten*) in Japanese Mahjong. Deficiency calculation is a basic procedure that can be used in all Mahjong variants by both rule-based and learning-based Mahjong AIs (see, e.g., [14], [6]). While algorithms for calculating deficiency or *shanten* exist (see Tenhou’s website for a *shanten* calculator and the quadtree method of [8]), these algorithms do not take the agent’s knowledge about available tiles into consideration, which makes their calculated numbers less useful.

The aim of this paper is to provide an efficient and general purpose algorithm for calculating the deficiency (*shanten*) number which respects the agent’s knowledge of available tiles. Recall that a hand is a winning hand if it can be decomposed into four melds (i.e., triples or three consecutive tiles) plus one pair.² We call any such combination a *decomposition*. As in the quadtree method [8], we represent the agent’s knowledge of available tiles as a k -tuple of integers

¹<https://botzone.org.cn/static/gamecontest2020a.html>

²There are possibly other kinds of winning hands in different Mahjong variants, but this is the most typical one.

less than or equal to 4, where k is the number of different tiles used in the Mahjong variant. Furthermore, we divide the hand into blocks of tiles, calculate their quasi-decompositions and map them into a small set of types, and then merge these local types and evaluate their costs, where the precise meaning of a quasi-decomposition will become clear in Definition 11. For the time being, we regard it as an incomplete version of decomposition. Meanwhile, the type of a quasi-decomposition is a 7-tuple which consists of values of seven attributes of the quasi-decomposition.

The remainder of this paper is structured as follows. After a short introduction of basic notions of Mahjong in Sec. II, we recall in Sec. III the quadtree method for calculating deficiency introduced in [8] and point out its limitations. The revised knowledge-aware definition of deficiency is introduced in Sec. IV, where we also introduce seven attributes of quasi-decompositions. Furthermore, we show that the revised deficiency can be obtained by calculating the minimal cost of all quasi-decompositions and the cost of a quasi-decomposition can be determined, in most time, by its attributes. Our block-based algorithm for calculating deficiency is described in Sec. V, where we show that if the deficiency is not greater than four, then our algorithm can return the number exactly. Experiments in Sec. VI demonstrate that the block-based algorithm is efficient and mostly exact. The last section concludes the paper with outlook on future research directions. Technical proofs can be found in the appendix.

II. BACKGROUNDS

In this section, we recall from [8] some basic notions of Mahjong. For simplicity, we only consider a very basic and essential version of Mahjong. The other variants can be dealt with analogously. There are three types of tiles used in the basic version of Mahjong:

- Bamboos: $B1, B2, \dots, B9$, each with four identical tiles
- Characters: $C1, C2, \dots, C9$, each with four identical tiles
- Dots: $D1, D2, \dots, D9$, each with four identical tiles

We call *Bamboo* (B), *Character* (C), *Dot* (D) *colours* of the tiles, and write \mathcal{M}_0 for the set of these tiles, which includes in total 108 tiles.

Definition 1 (pong, chow, meld). A *pong* is a sequence of three identical tiles. A *chow* is a sequence of three consecutive tiles of the same colour. A *meld* is either a pong or a chow.

The following non-standard notions are also useful.

Definition 2 (pchow, pmeld [8]). A *prechow* (*pchow*) is a pair of tiles of the same colour s.t. it becomes a chow if we add an appropriate tile with the same colour. A *premeld* (*pmeld*) is a pchow or a pair. We say a tile c *completes* a pmeld (ab) if (abc) (after reordering) is a meld. Similarly, a pair is completed from a single tile t if it is obtained by adding an identical tile to t .

For example, ($B3B4B5$) is a chow, ($C1C1$) is a pair, ($B7B7B7$) is a pong, and ($B1B3$) and ($C2C3$) are two pchows.

In a Mahjong game, each player begins with a set of 13 tiles (called her *hand*). Players may change a tile by either drawing a new tile or robbing a tile, say t , discarded by another player if the player has two identical tiles t and wants to consolidate a pong (ttt) or if t completes her 13-tile into a winning hand.

Definition 3 (hand). A k -tile is a sequence S of k tiles from \mathcal{M}_0 , where no identical tiles can appear more than four times. A hand is a 13- or 14-tile.

To be precise, we write a tile t as a pair (c, n) s.t. $c \in \{0, 1, 2\}$ denotes the colour of the tile t and n denotes the number of t . Here 0, 1, 2 denote, respectively, Bamboo, Character, and Dot. For example, (0, 3) and (1, 5) denote $B3$ and $C5$ respectively. In addition, we assume that each k -tile S is alphabetically ordered, e.g., (0, 3) is always before (1, 5) if both appear in S .

Suppose S is a sequence of k tiles. Write $S[i]$ for the $(i+1)$ -th value of S for $0 \leq i \leq k-1$, i.e., we assume, as in Python and many other programming languages, that $S[0]$ represents the first entity of S .

Definition 4 (winning or complete hand, decomposition, eye [8]). A 14-tile over \mathcal{M}_0 is *winning* or *complete* if it can be decomposed into four melds and one pair.³ Given a complete 14-tile H , a *decomposition* π of H is a sequence of five subsequences of H s.t. $\pi[4]$ is a pair and, for $0 \leq i \leq 3$, each $\pi[i]$ is a meld. If this is the case, we call $\pi[4]$ the *eye* of the decomposition.

The following 14-tile H is complete as it has a decomposition π .

$$H = (B1B2B2B3B3B4B7B7B7)(C1C1)(D4D5D6) \\ \pi = (B1B2B3)(B2B3B4)(B7B7B7)(D4D5D6)(C1C1).$$

Definition 5 (deficiency [8]). The *deficiency number* (or simply *deficiency*) of a 14-tile H is defined recursively:

- H has deficiency 0 if it is complete;
- In general, for $\ell \geq 0$, H has deficiency $\ell + 1$ if it has no deficiency smaller than or equal to ℓ and there exists a tile t in H and another tile t' s.t. $H[t/t']$ (the 14-tile obtained from H by replacing t with t') has deficiency ℓ .

If the deficiency of H is ℓ , we write $\text{dfncy}(H) = \ell$.

The above definition can also be extended to 13-tiles: A 13-tile H has deficiency 1 if there exists a tile t s.t. H becomes complete after adding t . In this case, H is called a ready hand or *tenpai* in Japanese Mahjong. In general, a 13-tile H has deficiency $\ell + 1$ for $\ell > 1$ if it does not have deficiency smaller than or equal to ℓ and it has deficiency ℓ after some proper tile change.

A 14-tile H is often not complete. The deficiency number of H measures how bad H is by counting the number of necessary tile changes to make it complete.

³For ease of presentation, we don't regard a hand with seven pairs as complete.

Proposition 1. *The deficiency of any 14-tile H is not greater than 6.*

For example, the following hand has deficiency 6.

$$H = (B3B5B6B9)(C2C2C3C6C9)(D1D1D2D5D8)$$

It is one basic requirement for a Mahjong AI to tell how bad a hand is when playing Mahjong. To tell if her hand H is winning, the agent only needs to check if $\text{dfncy}(H) = 0$. In general, after drawing a tile from the Wall, the agent needs to select a tile to discard if it does not complete her hand. This can be done by comparing, for each tile t in H , if she has the maximum chance of decreasing the deficiency after discarding t and drawing a tile from the Wall. Analogously, suppose another player discards a tile t and the agent has a pair (tt) in her hand. She needs to decide if she performs a pong of t . She may adopt a strategy in which she performs the pong only when the action helps to reduce (or does not increase) the hand deficiency.

III. THE QUADTREE ALGORITHM FOR DEFICIENCY COMPUTING

In this section we describe the quadtree algorithm for deficiency computing introduced in [8], which is based on calculating and comparing the costs of all pre-decompositions of a 14-tile H .

In this paper, we regard any sequence of tiles as a multiset, allowing for multiple instances for each of its elements. The usual set operations \cup, \cap, \setminus are also defined for multisets. For clarity, we also write a multiset, e.g., $\{B5, B5, B6, C8, C9\}$, as ordered tuples like $(B5B5B6)(C8C9)$.

A. Pre-decomposition

Given a 14-tile H , we can determine if $\text{dfncy}(H) \leq k$ by recursively checking if it has a neighbour $H[t/t']$ with deficiency $< k$, where $H[t/t']$ denotes the hand obtained by replacing a tile t in H by another tile t' . This is extremely inefficient. The more practical quadtree method is based on the notion of pre-decomposition.

Definition 6 (pre-decomposition and its remainder [8]). *A pre-decomposition (pDCMP) is a sequence π of five sequences, $\pi[0], \dots, \pi[4]$, s.t.*

- $\pi[4]$ is a pair, a single tile, or empty;
- for $0 \leq i \leq 3$, each $\pi[i]$ is a meld, a pmeld, a single tile, or empty.

We call each $\pi[i]$ for $0 \leq i \leq 3$ a meld holder and call $\pi[4]$ the eye holder. We say π is a complete pDCMP if it is a decomposition. If $\bigcup_{i=0}^4 \pi[i]$ is contained in a 14-tile H , we say π is a pDCMP of H and call the set of tiles in H that are not in $\bigcup_{i=0}^4 \pi[i]$ the *remainder* of H under π .

A decomposition of a complete hand H is also a pDCMP of H .

Definition 7 (completion and cost [8]). Suppose π and π' are two pDCMPs. We say π' is *finer* than π if, for each $0 \leq i \leq 4$, $\pi[i]$ is identical to or a subsequence of $\pi'[i]$. A pDCMP π is

completable if there exists a decomposition π^* that is finer than π . If this is the case, we call π^* a *completion* of π .

The *cost* of a completable pDCMP π , written $\text{cost}(\pi)$, is the number of missing tiles required to complete $\pi[4]$ into a pair and complete each $\pi[i]$ into a meld for $0 \leq i \leq 3$. If π is incompletable, we say it has infinite cost.

A 14-tile H may have many different pDCMPs.

Example 1. Consider the 14-tile

$$H = (B1B5B6B8B8B8B9)(\cdot)(D1D2D4D5D5D6D7)$$

It is easy to check that the following are all its pDCMPs

$$\pi_0 = (B5B6)(B8B8)(D4D5D6)(D5D7)(B8)$$

$$\pi_1 = (B5B6)(B8B8B8)(D4D5D6)(D5D7)(B1)$$

$$\pi_2 = (B5B6)(B8B8B8)(D4D5)(D5D6D7)(B1).$$

The meld holder $(B8B8)$ and the eye holder $(B8)$ in π_0 cannot be both completed, as there are only four $B8$ in Mahjong. The pDCMPs π_1 and π_2 have, respectively, the following completions

$$\pi_1^* = (\underline{B4B5B6})(B8B8B8)(D4D5D6)(D5D6D7)(\underline{B1B1})$$

$$\pi_2^* = (\underline{B4B5B6})(B8B8B8)(D4D5D6)(D5D6D7)(\underline{B1B1}).$$

We have $\text{cost}(\pi_1) = \text{cost}(\pi_2) = 3$. It is easy to check that $\text{dfncy}(T) = 3$. Therefore, both π_1 and π_2 are pDCMPs with the minimal cost.

Theorem 1 ([8]). *The deficiency of a 14-tile H is the minimum cost of its pDCMPs. We say a pDCMP π of a 14-tile H is minimal if $\text{cost}(\pi) = \text{dfncy}(T)$.*

We next present the deficiency calculation method introduced in [8].

B. The quadtree method

The quadtree method determines the deficiency of a 14-tile H by constructing and evaluating its possible pDCMPs in an exhaustive way. In the quadtree, each node is represented by a word α in the alphabet $\Sigma = \{1, 2, 3, 4\}$, where each symbol represents a possible action. Each node α is attached with a pDCMP π_α of H and a subsequence S_α of H , which denotes the set of tiles remaining to be processed. As will become clear, two words α, β are identical iff $(\pi_\alpha, S_\alpha) = (\pi_\beta, S_\beta)$. In what follows, we do not distinguish between a node α and its associated (π_α, S_α) . In the construction process, when S_α becomes empty, we terminate the branch with an exact cost estimation. Note that it is possible that π_α may still have empty placeholders and can be refined by recycling tiles from the remainder of π_α . For each empty placeholder, we decrease the cost by 1. This estimation is exact except a few exceptions that are treated specially (see [8]).

We use a queue Q to keep all nodes to be explored and a value val to denote the current minimum cost. The root node is denoted by the empty word ε , which has $S_\varepsilon = H$ and π_ε defined as $\pi_\varepsilon[i] = \emptyset$ for $0 \leq i \leq 4$. Initially, we set $\text{val} = 6$,

the largest possible deficiency number and put the root node ε in Q .

Suppose a node $\alpha = (S_\alpha, \pi_\alpha)$ is popped out from the queue Q . We *expand* α as follows. Let $t = S_\alpha[0]$ be the first tile in the subsequence S_α . We add up to four child nodes α_1 to α_4 under the node α . For each $\ell \in \{1, 2, 3, 4\}$, $S_{\alpha\ell}$ and $\pi_{\alpha\ell}$ are obtained by, respectively, *reducing* S_α and *refining* π_α as follows.

a) *Notation:* Given a tile $t = (c, n)$, we write t^+ and t^{++} for $(c, n+1)$ and $(c, n+2)$, respectively, as long as they are legal Mahjong tiles.

1. (pass) Define $S_{\alpha 1} = S_\alpha \setminus (t)$ and $\pi_{\alpha 1} = \pi_\alpha$.
2. (make chow) If t^+ or t^{++} is in S_α , define $S_{\alpha 2} = S_\alpha \setminus (tt^+t^{++})$. Suppose i is the first index in $\{0, 1, 2, 3\}$ s.t. $\pi_\alpha[i] = \emptyset$. Define $\pi_{\alpha 2}[i] = (tt^+t^{++}) \cap S_\alpha$ and $\pi_{\alpha 2}[j] = \pi_\alpha[j]$ for $j \neq i$.
3. (make eye) If $(tt) \subseteq S_\alpha$ and $\pi_\alpha[4]$ is empty, define $S_{\alpha 3} = S_\alpha \setminus (tt)$ and $\pi_{\alpha 3}[4] = (tt)$ and $\pi_{\alpha 3}[j] = \pi_\alpha[j]$ for $0 \leq j \leq 3$.
4. (make pong) If $(tt) \subseteq S_\alpha$, define $S_{\alpha 4} = S_\alpha \setminus (ttt)$. Suppose i is the first index in $\{0, 1, 2, 3\}$ s.t. $\pi_\alpha[i] = \emptyset$. Define $\pi_{\alpha 4}[i] = (ttt) \cap S_\alpha$ and $\pi_{\alpha 4}[j] = \pi_\alpha[j]$ for $j \neq i$.

If a certain condition is not satisfied, e.g., when neither t^+ nor t^{++} is in S_α , this node is not introduced. For each child node $\alpha\ell$, if $\pi_{\alpha\ell}[i] \neq \emptyset$ for all $0 \leq i \leq 4$, we compare the value $\text{cost}(\pi_{\alpha\ell})$ with the current best value val , update val as $\text{cost}(\pi_{\alpha\ell})$ if the latter is smaller, and terminate this branch.

Suppose $\pi_{\alpha\ell}[i]$ is empty for some i . If there are tiles left to process, i.e., $S_{\alpha\ell} \neq \emptyset$, we put $\alpha\ell$ in Q and expand it later; otherwise, we derive the cost of a refinement of $\pi_{\alpha\ell}$ that has no empty placeholder (see [8] for details) and compare that cost with val , update val accordingly, and terminate this branch. If Q is nonempty and $\text{val} > 0$, we pop out the next node β from Q and expand β as above. The procedure terminates when $\text{val} = 0$ or $Q = \emptyset$.

Example 2. Consider the following 14-tile

$$H = (B1B4B7)(C2C5C8)(D1D4D7D8D8D9D9D9).$$

Let \mathcal{T} be the quadtree of the 14-tile H . For $1 \leq k \leq 8$, $\alpha_k = 1^k \equiv \underbrace{1 \cdots 1}_k$ is the only depth k node of \mathcal{T} that are not closed.

In particular, $\pi_{\alpha_8} = \pi_\varepsilon$ and $S_{\alpha_8} = (D7D8D8D9D9D9)$. As there are 6 tiles to be processed, we can make at most two melds or one meld and one eye, which left three places of the pDCMP empty. This implies that, for any node β that is a descendant of $\alpha_9 = 1^9$, we have $\text{cost}(\pi_\beta) \geq 6 \geq \text{val}$. Thus, it is not necessary to expand α_9 .

Consider the 14-tile and pDCMPs in Example 1 again. None of these pDCMPs is associated with a node in the quadtree, but π_0 is a direct refinement of π_α with $\alpha = 121411122$ by putting $B8$ in the eye holder. Analogously, π_1 is a direct refinement of π_β with $\beta = 12411122$ by putting $B1$ in the eye holder. However, even if we removing $B1$ from the eye holder, π_2 is still not associated with any node in the quadtree. This is because, when $S_\gamma = (D4D5D5D6D7)$

and we examine the tile $D4$, if we select action 2 and put $D4$ and $D5$ in a meld holder, then according to the quadtree construction we should also put $D6$ together with $D4$ and $D5$. This does not affect the correctness of the method as long as the current knowledge of available tiles is not concerned (cf. Example 3 below).

In the worst case, the quadtree has a maximum depth of 14, as S_{α_k} with $\alpha_k = 1^k$ is a node in the quadtree with length k ($0 \leq k \leq 14$). Often, we may stop earlier if we know for certain that no refinement of π_α can have a cost smaller than val .

Since this exhaustive method is inefficient, in practice, we have to adopt an approximate method by, e.g., considering only nodes with depth smaller than or equal to 8. While this is sufficient in most cases, it is still not very efficient and, in some cases, is inexact and may lead to bad decisions.

IV. KNOWLEDGE-AWARE DEFICIENCY

Another problem with the quadtree method is that it does not take into consideration the agent's knowledge about the game state. In this section, we propose a modified definition of the deficiency number, which respects the agent's knowledge about the available tiles in the current state of the game.

A. Knowledge base and knowledge-aware deficiency

The agent's knowledge base contains all her information of available tiles.

Definition 8 (knowledge base [8]). The knowledge base of the agent is a 27-tuple KB . For each tile $t = (c, n)$ ($0 \leq c \leq 2$, $1 \leq n \leq 9$), $KB(t) = KB(c, n) \triangleq KB[9c + n - 1]$ denotes the multiplicity of t the agent *believes* to be available.

In what follows, we say KB has a tile t , or say t is in KB , if $KB(t) > 0$. Similarly, we say KB has a pair (pong) if there is a tile t s.t. $KB(t) \geq 2$ ($KB(t) \geq 3$) and say KB has a chow if there is a chow (t_1, t_2, t_3) s.t. $\min(KB(t_1), KB(t_2), KB(t_3)) \geq 1$.

Initially, we have $KB(t) = 4$ for each tile t . When all players have their hands, the agent also has her hand H and updates her KB accordingly as

$$KB(t) = 4 - \text{the number of } t \text{ in } H. \quad (1)$$

Then she continues to modify KB according to the process of the game. For example, if one player discards a tile t and a pong of t is made by some other player, then the agent updates her KB decreasing by 3 her $KB(t)$ and leaves the other items unchanged. The agent may also update her KB by reasoning. For example, suppose player A declared win after robbing a tile $C2$. If the agent's KB shows that $C2$ and $C4$ are not available, she can infer that player A must have $C1$ and $C3$ in his hand (to make a meld $(C1C2C3)$). Accordingly, the agent could update her KB decreasing by one her $KB(C1)$ and $KB(C3)$.

In what follows, whenever a k -tile H and a knowledge base KB appear together, we always assume that they are

compatible in the sense that the number of any tile in H and KB is not greater than 4.

If every tile that can complete a hand H has been discarded, ponged, or robbed, H becomes incompletable in the current KB . This suggests that Definition 5 should take KB into consideration.

For convenience, for a k -tile H , a knowledge base KB , and any tile t , we write $H \oplus t$ as the $(k+1)$ -tile obtained by adding t to H (after reordering), and $KB \ominus t$ as the updated knowledge base which differs from KB only in that it has one less t than KB does.

Definition 9 (knowledge-aware deficiency). The *deficiency number* (or *deficiency*) of a 14-tile H w.r.t. a knowledge base KB is defined recursively:

- H has deficiency 0 if it is complete;
- In general, for $\ell \geq 0$, H has deficiency $\ell + 1$ if it has no deficiency smaller than or equal to ℓ and there exists a tile t in H and another tile t' that is available in KB (i.e., $KB(t') > 0$) s.t. $H[t/t']$ has deficiency ℓ w.r.t. the updated knowledge base $KB \ominus t'$.

If the deficiency of H is ℓ , we write $\text{dfncy}(H, KB) = \ell$. We say H is incompletable if it has no finite deficiency.

Suppose H is a 13-tile. The deficiency of H w.r.t. a knowledge base KB is defined as the minimum of $\text{dfncy}(H \oplus t, KB \ominus t) + 1$ over all tiles t with $KB(t) > 0$.

It is easy to see that, when ignoring the knowledge base KB , i.e., interpreting KB as Eq. (1), then $\text{dfncy}(T) = \text{dfncy}(T, KB)$.

Example 3. Consider the 14-tile H and its knowledge base KB

$$H = (B1B5B6B8B8B8B9)(\cdot)(D1D2D4D5D5D6D7),$$

$$KB = (343423023)(434434443)(334220344).$$

For clarity, we group the 27-tuple KB into three 9-tuples, representing the local knowledge bases of colour B, C, D . In the quadtree search process, the pDCMP at node $\alpha = (12311222)$ is $\pi = (B5B6)(D1D2)(D4D5D6)(D5D7)(B8B8)$. If we don't consider the knowledge base, π has cost 3. Let $\pi^* = (B5B6)(D1D2)(D4D5)(D5D6D7)(B8B8)$. Then π^* is another pDCMP with cost 3. However, π^* is not the pDCMP of any node in the quadtree! This is because, by construction, if $(D4D5D6)$ is a subsequence of S_β and $D4 = S_\beta[0]$, then we shall enrich π_β with the chow $(D4D5D6)$ instead of a pchow $(D4D5)$.

Taking KB into consideration, π is incompletable under KB as it has a pmeld $(D5D7)$ but $D6$ is not available in KB . In comparison, π^* is still completable under KB and has cost 3 in the sense of Definition 7.

This example shows that the quadtree method is incomplete when the knowledge base is taken into consideration, as some pDCMPs with the minimal cost are not reachable. In order to get a complete method, we could expand the quadtree into a 6-ary tree, where each node have up to six child nodes s.t. pmelds

like $(D4D5)$ and $(D4D6)$ could appear in the 6-ary tree of the above example. However, this will make the algorithm even slower (cf. Sec. VI for evaluation). Therefore, it is necessary to devise a new deficiency calculation algorithm which is efficient and respects the knowledge base. To this end, we introduce the following adapted notion of quasi-decompositions and reduce the deficiency of a hand w.r.t. to a knowledge base to the minimal cost over all its quasi-decompositions.

B. Quasi-decompositions

We need the following notion of completable pmeld.

Definition 10 (completable pmeld). A pchow (t, t') is completable under KB if KB has a tile t'' s.t. (t, t', t'') (after reordering) is a chow. Similarly, a pair (t, t) is completable under KB if KB has an identical copy of t (i.e., $KB(t) > 0$). A pmeld is completable if it is a completable pchow or a completable pair.

A quasi-decomposition is a variant of pre-decomposition that respects the knowledge base but does not fix the last place as the eye holder.

Definition 11 (quasi-decomposition). Let H be a k -tile and KB a knowledge base. A quasi-decomposition (qDCMP) $\pi = (\pi[0], \pi[1], \dots, \pi[k])$ of H w.r.t. KB is a set of (possibly identical) subsequences of H s.t.

- $k \leq 4$ and each $\pi[i]$ is a meld, a pair, or a pchow.
- If $k = 4$, π contains at least one pair.
- Except at most one pair, all pmelds in π are completable under KB .
- $\bigcup_{i=0}^k \pi[i]$ is contained in H .

The *remainder* of π in H is the sequence of tiles in H that are not in $\bigcup_{i=0}^k \pi[i]$.

Example 4. Let

$$H = (\cdot)(C1C4C6C7C8C9)(D1D2D3D6D6D7D8)$$

$$KB = (001100121)(010000030)(032242321)$$

As the knowledge base contains only one $C2$ and three $C8$, pchows $(C4C6)$, $(C6C8)$, $(C7C8)$, and $(C8C9)$ are incompletable. Then

$$\pi_0 = (\cdot),$$

$$\pi_1 = ((C6C7), (D1D2D3), (D6D6), (D7D8)),$$

$$\pi_2 = ((C7C9), (D1D2D3), (D6D6), (D7D8)),$$

$$\pi_3 = ((C6C7C8), (D1D2D3), (D6D6), (D7D8)),$$

$$\pi_4 = ((C7C8C9), (D1D2D3), (D6D6), (D7D8))$$

are examples of valid qDCMPs of H . In addition, although $(C6C7)$ and $(C7C9)$ are completable, they can only be completed by $C8$, which is also in the remainder of the corresponding qDCMPs, viz. π_1 and π_2 . Intuitively, their completion costs are larger than that of π_3 and π_4 and thus it is not necessary to evaluate them. In practice, we need only evaluate π_3 and π_4 .

Suppose π is a qDCMP of H under KB and R the remainder of π in H . The cost of π under KB , written $\text{cost}(\pi, KB)$, is the minimal number of tile changes to complete π into a decomposition by applying the following procedure until we are certain that π is incompletable or have four melds and one eye.

- (a) If π has an incompletable pair, select it as the eye; otherwise, select a pair in π to act as the eye or create the eye from scratch.
- (b) Complete every pmeld in π that does not act as the eye.
- (c) If by far π has three or less melds, create a new meld from scratch until π has four melds.

More precisely, suppose (t, t') is a pmeld in π which we want to complete. If (t, t') is a pchow, we borrow one tile from KB and complete it into a chow, increasing the cost by 1. Suppose (t, t) is a pair that is not selected as the eye in Step (a). Since it is completable, we borrow one tile t from KB (and increase the cost by 1) to complete (t, t) into a pong.

To create the eye from scratch, we first examine if there is a tile t in R which has an identical tile in KB .⁴ If so, we use (t, t) as the eye and this incurs cost 1. Suppose otherwise. We check if KB has a pair; if so, we use it as the eye with the cost increased by 2. In case that KB contains no pairs, we cannot create a new pair and need to go back to Step (a) and check if we can use a pair in π as the eye. If that is still impossible, then π is incompletable.

Analogously, to create a new meld from scratch, we increase the cost by 2 if it is possible to make a meld by using one tile from R and borrowing two tiles from KB ; otherwise, we increase the cost by 3 if KB contains a meld. If neither is true, then π is incompletable.

Suppose the procedure terminates with a decomposition π^* and the total cost is c . We call π^* a completion of π and c the cost of π , written as $\text{cost}(\pi, KB)$.

Similar to Theorem 1, knowledge-aware deficiency has the following characterisation.

Theorem 2. *Let H be a hand and KB a knowledge base. The deficiency of H w.r.t. KB is the minimal of $\text{cost}(\pi, KB)$ for all qDCMPs π of H under KB .*

The next subsection shows that the cost of a qDCMP can be determined by its attributes.

C. Attributes of a qDCMP

Let H be a k -tile and KB a knowledge base. Suppose π is a qDCMP of H and R the remainder of π in H . We introduce the following attributes for π , where m, n, p, e are non-negative integers and re, rm, em are Boolean values.

- m The number of melds in π .
- n The number of pmelds in π , including pchows and pairs, where each pchow is completable w.r.t. KB .
- p The number of pairs in π .

⁴It is possible that R contains a pair (t, t) , we don't regard the cost to create the eye as 0. This is because the deficiency is determined by considering all qDCMPs and (t, t) appears in some other qDCMP π' that is finer than π and has a smaller cost.

- e The number of pairs in π which are incompletable.
- re If the remainder has a tile t which is also in KB , then $re = 1$; otherwise, it's 0. If $re = 1$, we can make the eye based on one tile recycled from R .
- rm If the remainder has a tile t which can evolve into a meld (i.e., a chow or a pong) given the current KB , then $rm = 1$; otherwise, it's 0. If $rm = 1$, we can make a meld based on one tile recycled from R .
- em Suppose $e = 0$ and $re = rm = 1$ and we need to make a meld and the eye both from scratch. If we cannot make a meld from a tile in R after making the eye starting from a tile in R , or vice versa, then we say there is an *eye-meld conflict* and set $em = 1$. In all other cases, we set $em = 0$.

In addition, we have the following attributes for KB .

- ke If the KB contains a pair, then $ke = 1$; otherwise, it's 0.
- km If the KB contains a meld, then $km = 1$; otherwise, it's 0.

a) *About the notations:* The notation xy for $x \in \{r, k\}$ and $y \in \{e, m\}$ reads as, if $y = e$ ($y = m$, resp.), we can make the **eye** (a **meld**, resp.) from r (the **remainder**) or k (the **knowledge base**).

The following result is clear from the definition.

Lemma 1. *Let H be a k -tile, KB a knowledge base, and π a qDCMP of H under KB . Suppose m, n, p, e are attributes of π defined above. Then $e \leq 1$, $m + n \leq 5$, $3m + 2n \leq k$, and $n \geq p \geq e \geq 0$. If $m + n = 5$, then $p > 0$.*

Definition 12 (type of qDCMP). Let H be a k -tile and KB a knowledge base. Suppose π is a qDCMP of H under KB . Let m, n, p, e, re, rm, em be attributes of π defined as above. We call $\sigma(\pi) \triangleq (m, n, p, e, re, rm, em)$ the type of π .

We next show how the cost of π under KB can be determined by its type.

Lemma 2. *Let π be a qDCMP of H under KB . Suppose $\sigma(\pi) = (m, n, p, e, re, rm, em)$ is the type of π and ke, km the two attributes of KB . Then π is incompletable if*

- (a) $p = re = ke = 0$; or
- (b) $m + n \leq 4$, $rm = km = 0$, and $re = ke = 0$; or
- (c) $m + n - e \leq 3$ and $rm = km = 0$; or
- (d) $m + n \leq 3$, $p = ke = km = 0$ and $re = rm = em = 1$.

The above conditions are sufficient but not necessary. However, we can easily see from the following two lemmas that they are necessary when $m + n \geq 4$, or when $m + n = 3$ and $p = 0$.

We next examine the cases when $m + n \geq 4$ and $m + n - e = 3$.

Lemma 3. *Let π be a qDCMP of H under KB and (m, n, p, e, re, rm, em) its type. Suppose $m + n \geq 4$.*

- (a) *If $m + n > 4$, then the cost of π is $4 - m$.*
- (b) *Suppose $m + n = 4$. If $e = 0$ and $re = 1$, or $p > 0$ and $rm = 1$, then the cost of π is $4 - m + 1$.*

- (c) Suppose $m + n = 4$, $re = 0$ if $e = 0$, and $rm = 0$ if $p > 0$.⁵ If $e = 0$ and $ke = 1$, or $p > 0$ and $km = 1$, then the cost of π is $4 - m + 2$.
- (d) If none of the above, then π is incompletable.

Note in this case the eye-meld conflict does not matter as we do not need to make the eye and a new meld simultaneously. The next lemma considers the cases when the number of melds and completable pmelds in π is three.

Lemma 4. Let π be a qDCMP of H under KB and (m, n, p, e, re, rm, em) its type. Suppose $m + n - e = 3$ and $\max(p, re, ke) > 0$ and $\max(rm, km) = 1$. Let $mcost = 2 \times rm + 3 \times (1 - rm)$ and $ecost = re + 2 \times (1 - re)$.

- (a) If $e = 1$, then the cost of π is $n - 1 + mcost$.
- (b) If $p = em = 0$, then the cost of π is $n + mcost + ecost$.
- (c) If $p = 0$, $em = 1$, and $\max(ke, km) = 1$, then the cost of π is $n + 4 \geq 4$.
- (d) If $p > e = 0$, then the cost of π is at least $\min(f_1, f_2) \geq 4$, where $f_1 \triangleq n + mcost + ecost$ and $f_2 \triangleq n - 1 + 2 \times mcost$.

When $m + n - e \leq 2$, we can analyse as above. In order to get the exact cost, we need to introduce new attributes and consider many more subcases. This is in general not necessary as the cost of a qDCMP in this case is at least 4, which is quite bad and usually of little use as it indicates that either (i) the qDCMP is not a good one and may be replaced with a better one, or (ii) the hand is pretty bad, and, if this is in the late stage of the game, we may have no chance to complete the hand.

Lemma 5. Let π be a qDCMP of H under KB and (m, n, p, e, re, rm, em) the type of π . If $m = 0$ or $m + n - e \leq 2$, then the cost of π is at least 4.

When $m + n - e \leq 2$, similar to Lemma 4, we have the following result.

Lemma 6. Let π be a qDCMP of H under KB and (m, n, p, e, re, rm, em) the type of π . Suppose $m + n - e \leq 2$. Let $mcost = 2 \times rm + 3 \times (1 - rm)$ and $ecost = re + 2 \times (1 - re)$.

- (a) Suppose $e = 1$. The cost of π is at least $n - 1 + mcost \times (4 - m - n + 1) \geq 4$.
- (b) Suppose $p = 0$. The cost of π is at least $n + mcost \times (4 - m - n) + ecost + em \geq 5$.
- (c) If $p > e = 0$, then the cost of π is at least $\min(f_1, f_2) \geq 6$, where $f_1 \triangleq n + mcost \times (4 - n - m) + ecost$ and $f_2 \triangleq n - 1 + mcost \times (4 - m - n + 1)$.

Example 4 (cont'd). Recall

$$H = (\cdot)(C1C4C6C7C8C9)(D1D2D3D6D6D7D8) \\ KB = (001100121)(010000030)(032242321).$$

As KB contains pairs and melds, we have $ke = km = 1$. Consider the qDCMP $\pi \triangleq (C6C7C8)(D1D2D3)(D6D7D8)$

⁵Here a statement ' B if A ' is equivalent to saying ' $A \Rightarrow B$ ' or ' $(\text{not } A) \text{ or } B$ '.

with remainder $R \triangleq (C1C4C9D6)$. It is easy to check that π has type $(3, 0, 0, 0, 1, 1, 1)$. Since $ke = km = 1$, by Lemma 4 (c), the cost of π is 4. After examining all qDCMPs of H , we know this is the minimal cost and thus H has deficiency 4 under KB , i.e., $\text{dfncy}(H, KB) = 4$.

We integrate the above results in Alg. 1 to determine the cost of a qDCMP.

Algorithmus 1 *Decide*($m, n, p, e, re, rm, em, ke, km$)

Require: A qDCMP type $\vec{x} = (m, n, p, e, re, rm, em)$ and ke, km .
Ensure: The cost of \vec{x} .

```

1: if  $p = re = ke = 0$  then return 100 end if
2: if  $m + n \leq 4$  and  $re = ke = rm = km = 0$  then return 100
   end if
3: if  $m + n - e \leq 3$  and  $rm = km = 0$  then return 100 end if
4: if  $m + n \leq 3$ ,  $p = ke = km = 0$ , and  $em = 1$  then return 100
   end if
5: if  $m + n \geq 4$  then
6:   if  $m + n > 4$  then
7:     return  $4 - m$ 
8:   else
9:     if  $(e = 0 \text{ and } re = 1) \text{ or } (p > 0 \text{ and } rm = 1)$  then
10:      return  $4 - m + 1$ 
11:    else
12:      return  $4 - m + 2$ 
13:    end if
14:  end if
15: end if
16: if  $m + n - e \leq 3$  then
17:    $mcost \leftarrow 2 \times rm + 3 \times (1 - rm)$ 
18:    $ecost \leftarrow re + 2 \times (1 - re)$ 
19:   if  $e = 1$  then
20:     return  $(n - 1) + mcost \times (4 - m - n + 1)$ 
21:   else
22:     if  $p = 0$  then  $\triangleright p = e = 0$ 
23:       return  $n + mcost \times (4 - m - n) + ecost + em$ 
24:     else  $\triangleright p > e = 0$ 
25:        $f_1 \leftarrow n + mcost \times (4 - m - n) + ecost$ 
26:        $f_2 \leftarrow n - 1 + mcost \times (4 - m - n + 1)$ 
27:       return  $\min(f_1, f_2)$ 
28:     end if
29:   end if
30: end if

```

V. THE BLOCK DEFICIENCY ALGORITHM

In the quadtree algorithm, the hand H is considered as a whole and we generate, in principle, all possible pDCMPs of H and evaluate their costs one by one. This procedure can be significantly sped up by dividing the hand into blocks, generating the local quasi-decompositions and mapping them to small sets of local types and then merging these local types and evaluating their costs. To describe this block-based algorithm, we first introduce the notion of block.

Definition 13 (block). A block of a k -tile H is a subsequence b of H s.t.

- $b \neq \emptyset$ and all tiles in b have the same colour;
- If t is a tile in b , then any tile in H that is connected to t is also in b ,

where two tiles $t = (c, n)$, $t' = (c', n')$ are connected if $c = c'$ and $|n - n'| \leq 2$.

Analogously to deficiency, blocks can also be knowledge-aware.

Definition 14 (knowledge-aware block). Given a k -tile H and a knowledge base KB , a subsequence b of H is a KB -block if

- $b \neq \emptyset$ and all tiles in b have the same colour;
- If t is a tile in b , then any tile in H that is KB -connected to t is also in b ,

where two tiles $t = (c, n)$ and $t' = (c', n')$ are KB -connected if (i) $t = t'$, or (ii) either H or KB has a tile t'' s.t. $(tt't'')$ is a chow.

It is easy to see that there are blocks that are not KB -blocks and vice versa. In this paper, we are mainly concerned with KB -blocks.

In this section, we use Example 4 as our running example.

Example 4 (cont'd). Clearly, H has the following KB -blocks

$$b_1 \triangleq (C1), b_2 \triangleq (C4), b_3 \triangleq (C6C7C8C9), \\ b_4 \triangleq (D1D2D3), b_5 \triangleq (D6D6D7D8).$$

We note that $C4$ is not KB -connected to $C6$ as neither H nor KB contains a tile $C5$. All tiles in a KB -block, say $b_3 = (C6C7C8C9)$, have the same colour. For simplicity, we write it as a list of integers, say (6789), and write, say, $KB_1 = (010000030)$ for the restriction of KB to the Character suit that contains b .

The first block $b_1 = (C1)$ has only the empty qDCMP $\pi_1 = (\cdot)$, with remainder $R_1 = (C1)$ and local knowledge base $KB_1 = (010000030)$. As we cannot develop $C1$ into a pair or a meld, we have $re_1 = rm_1 = em_1 = 0$ and thus $\sigma(\pi_1) = (0, 0, 0, 0, 0, 0, 0)$. The second block $b_2 = (C4)$ is similar. It has the empty qDCMP π_2 with type $(0, 0, 0, 0, 0, 0, 0)$. The third block $b_3 = (C6C7C8C9)$ also has the local knowledge base KB_1 in the Character suit. For the qDCMP $\pi_3 = ((C6C7C8))$ of b_3 , as its remainder is $(C9)$, we have $m_3 = 1, n_3 = p_3 = e_3 = 0$, and $re_3 = rm_3 = em_3 = 0$. Thus the type of π_3 is $(1, 0, 0, 0, 0, 0, 0)$. Furthermore, consider the blocks $b_4 = (D1D2D3)$ and $b_5 = (D6D6D7D8)$ in the Dot suit, with the local knowledge base $KB_2 = (032242321)$. Clearly, $\pi_4 = ((D1D2D3))$ and $\pi_5 = ((D6D7D8))$ are, respectively, qDCMPs of b_4 and b_5 under KB_2 . The type of π_4 is $(1, 0, 0, 0, 0, 0, 0)$. Let $(m_5, n_5, p_5, e_5, re_5, rm_5, em_5)$ be the type of π_5 . As the remainder of π_5 is $(D6)$, we have $m_5 = 2, n_5 = p_5 = e_5 = 0, re_5 = rm_5 = 1$. Since the remainder contains a single tile, we cannot make the eye and a meld from the remainder $(D6)$ simultaneously, i.e., we have an eye-meld conflict and $em_5 = 1$. Thus the type of π_5 is $(2, 0, 0, 0, 1, 1, 1)$.

A. Description of the new algorithm

The idea is to first partition the hand H into pairwise disconnected KB -blocks. Write \mathcal{B} for the set of KB -blocks of H . For each KB -block b , we generate all its qDCMPs and call them *local* qDCMPs at b . Instead of amalgamating these local qDCMPs directly (and form global qDCMPs), we calculate

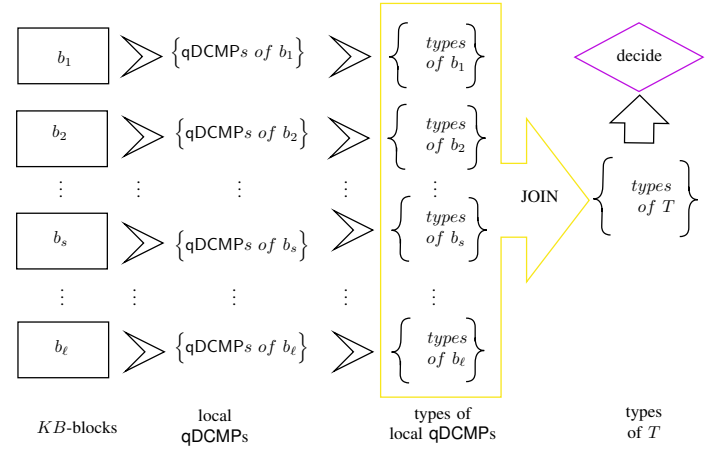


Fig. 1. Illustration of the block deficiency algorithm.

and join their *types*. Here the type of a local qDCMP of KB -block b is a 7-tuple $\vec{x}_b \triangleq (m_b, n_b, p_b, e_b, re_b, rm_b, em_b)$ (cf. Definition 12). Write $TypeSet_b$ for the set of types of all local qDCMPs of block b . We join these local types as follows. Let \vec{x}_b be a type in $TypeSet_b$ for each KB -block b . The global type $\vec{x} \triangleq (m, n, p, e, re, rm, em)$ is defined by setting

$$em = 1 \text{ iff } (\exists b \in \mathcal{B}) \\ (em_b = 1 \wedge (\forall b' \in \mathcal{B})(b \neq b' \rightarrow re_{b'} = rm_{b'} = 0))$$

and setting the other attributes as follows:

$$\vec{x}[j] = \begin{cases} \sum_b \vec{x}_b[j] & \text{if } 0 \leq j \leq 3 \\ \max_b \vec{x}_b[j] & \text{if } 4 \leq j \leq 5 \end{cases}$$

Note that in practice we join these local type sets one by one (see Alg. 3), as the combined type set usually have size far less than the product of the sizes of two local type sets. For each global type \vec{x} , we determine its cost according to Alg. 1, that is, the minimum number of tiles we should borrow from the knowledge base in order to complete the corresponding quasi-decomposition. Note that we ignore a global type $\vec{x} = (m, n, p, e, re, rm, em)$ if $e > 1$ or $m+n > 5$, as no legal qDCMP has type like that (cf. Lemma 1). Taking the minimum of the costs over all global types, we obtain the deficiency of the hand H .

Alg. 2 presents the pseudocode of the block deficiency algorithm.

B. The correctness of Alg. 2

To show that Alg. 2 is correct, we need to show that the type set of all global qDCMPs of H can be obtained by joining types of all local qDCMPs.

Definition 15 (Restriction of qDCMPs). Suppose H is a hand and KB a knowledge base. Let b be a KB -block of H . For any qDCMP π of H , the restriction of π to b , written $\pi_{\downarrow b}$, is the subset of π that includes all melds and pmelds of π that are contained in b .

Algorithmus 2 The block deficiency algorithm

Require: A 14-tile H and a knowledge base KB .**Ensure:** The deficiency of H .

```
1:  $\mathcal{B} \leftarrow KB\text{-blocks of } H$ 
2:  $\triangleright$  Generate the local type set for each  $KB$ -block  $b$ 
3: for  $b$  in  $\mathcal{B}$  do
4:    $TypeSet_b \leftarrow \emptyset$ 
5:    $DCMP_b \leftarrow$  all qDCMPs of  $b$ 
6:    $\triangleright$  Analogous to the quadtree method
7:   for  $\pi_b \in DCMP_b$  do
8:     Add the type of  $\pi_b$  to  $TypeSet_b$ 
9:   end for
10: end for
11:  $\triangleright$  Combine the local type sets and get the global type set
12:  $TypeSet \leftarrow \emptyset$ 
13: for  $b \in \mathcal{B}$  do
14:   if  $TypeSet = \emptyset$  then
15:      $TypeSet \leftarrow TypeSet_b$ 
16:   else
17:      $TypeSet \leftarrow TypeSet \bowtie TypeSet_b$   $\triangleright$  See Alg. 3
18:   end if
19: end for
20:  $\triangleright$  Evaluate each global type and obtain the deficiency
21:  $dfncy \leftarrow 100$   $\triangleright 100$  is a large upperbound
22:  $ke, km \leftarrow 0, 0$ 
23: if  $KB$  has a pair then  $ke \leftarrow 1$  end if
24: if  $KB$  has a meld then  $km \leftarrow 1$  end if
25: for  $\vec{x} = (m, n, p, e, re, rm, em) \in TypeSet$  do
26:    $dfncy' \leftarrow Decide(m, n, p, e, re, rm, em, ke, km)$ 
27:    $\triangleright$  See Alg. 1
28:   if  $dfncy' = 0$  then return 0 end if
29:    $dfncy \leftarrow \min(dfncy, dfncy')$ 
30: end for
31: return  $dfncy$ 
```

Algorithmus 3 Combine two type sets: $TypeSet \bowtie TypeSet'$

Require: Two type sets $TypeSet$ and $TypeSet'$.**Ensure:** The combined type set $TypeSetX$.

```
1:  $TypeSetX \leftarrow \emptyset$ 
2: for  $\vec{x} = (m, n, p, e, re, rm, em) \in TypeSet$  do
3:   for  $\vec{y} = (m', n', p', e', re', rm', em') \in TypeSet'$  do
4:      $m, n, p, e \leftarrow m + m', n + n', p + p', e + e'$ 
5:     if  $e > 1$  or  $m + n > 5$  or  $(m + n = 5 \text{ and } p = 0)$  then
6:       Continue
7:     else
8:       if  $(em = 1, re' = rm' = 0)$  or  $(em' = 1, re = rm = 0)$  then
9:          $em \leftarrow 1$ 
10:       else
11:          $em \leftarrow 0$ 
12:       end if
13:       Add  $(m, n, p, e, \max(re, re'), \max(rm, rm'), em)$ 
14:       to  $TypeSetX$ 
15:     end if
16:   end for
17: end for
18: Return  $TypeSetX$ 
```

Any restriction of a qDCMP to a KB -block is also a qDCMP. On the other hand, qDCMPs on different KB -blocks can be amalgamated.

Definition 16 (Amalgamation of local qDCMPs). Suppose H is a hand and KB a knowledge base. Let \mathcal{B} be the set of KB -blocks of H . For each $b \in \mathcal{B}$, let π_b be a local qDCMP on b . The amalgamation $\Pi_{b \in \mathcal{B}} \pi_b$ of all π_b is the union of all π_b .

The amalgamation of all π_b is a qDCMP on H if (i) it has at most five elements and at most of one of which is an incompletable pair; and (ii) it has at least one pair when it has five elements.

The following lemma shows that any qDCMP of a hand H can be recovered from its restrictions to the KB -blocks of H .

Lemma 7. Let \mathcal{B} be the KB -blocks of H . Suppose π is a qDCMP of H . For any $b \in \mathcal{B}$, let $\pi_{\downarrow b}$ be the restriction of π to b . Then $\pi_{\downarrow b}$ is a qDCMP of b . Moreover, π is exactly the amalgamation of all $\pi_{\downarrow b}$, i.e., $\pi = \Pi_{b \in \mathcal{B}} \pi_{\downarrow b}$.

In addition, the type of a global qDCMP is also uniquely determined by the types of its corresponding local qDCMPs.

Lemma 8. Suppose H is a hand and π a qDCMP on H . Let $\vec{x} = (m, n, p, e, re, rm, em)$ be the type of π . For each KB -block b of H , let $\vec{x}_b = (m_b, n_b, p_b, e_b, re_b, rm_b, em_b)$ be the type of $\pi_{\downarrow b}$, the restriction of π to b . Then we have

- $em = 1$ iff $e = 0$ and there exists $b \in \mathcal{B}$ s.t. $em_b = 1$ and $re_{b'} = rm_{b'} = em_{b'} = 0$ for any $b' \neq b$.
- $\vec{x}[i] = \sum_b \vec{x}_b[i]$ for $0 \leq i \leq 3$ and $\vec{x}[i] = \max_b \vec{x}_b[i]$ for $4 \leq i \leq 5$.

Example 4 (cont'd). Joining the four local qDCMPs π_i ($1 \leq i \leq 4$) we have a qDCMP for the hand H : $\pi = (C6C7C8)(D1D2D3)(D6D7D8)$, with remainder $(C1C4C9D6)$. Let $\sigma(\pi_i) = (m_i, n_i, p_i, e_i, re_i, rm_i, em_i)$ be the type of π_i for $1 \leq i \leq 5$. These are $(0, 0, 0, 0, 0, 0, 0)$, $(0, 0, 0, 0, 0, 0, 0)$, $(1, 0, 0, 0, 0, 0, 0)$, $(1, 0, 0, 0, 0, 0, 0)$ and $(1, 0, 0, 0, 1, 1, 1)$. Thus, the joint qDCMP has type $(3, 0, 0, 0, 1, 1, 1)$, which is identical to the result obtained by using Lemma 8.

The correctness of Alg. 2 then follows directly from the above lemmas.

Theorem 3. Suppose H is a hand and KB a knowledge base. Let $block_dfncy$ be the number returned by Alg. 2. Then $block_dfncy \leq dfncy(H, KB)$. Moreover, if $dfncy(H, KB) \leq 4$, then $block_dfncy = dfncy(H, KB)$.

Remark. In above we assume that the hand is a 13- or 14-tile. In practice, a player may also have $0 \leq k \leq 4$ consolidated melds (chows, pongs, or kongs) and her hand is then a $(13 - 3k)$ - or $(14 - 3k)$ -tile. When calculating the type of any qDCMP of H , we need to increase its m -index by k . The calculation procedure is similar to the cases when H is a 13- or 14-tile.

TABLE I
EXPERIMENTAL RESULTS ON PURE HANDS

num. tiles in KB	num. hand/KB pairs	quadtree max. calc. time (ms)	block max. calc. time (ms)	quadtree avg. calc. time (ms)	block avg. calc. time (ms)	avg. ratio
[0, 5)	4767	1279.2	78.2	86.3	13.1	6.6
[5, 10)	8923	5133.3	110.6	604.8	29.6	20.4
[10, 15)	9635	7280.3	112.7	1319.3	40.1	32.9
[15, 22]	6675	9344.9	115.8	2197.5	44.8	49.0
[0,22]	30,000	9344.9	115.8	1106.3	33.7	32.8

VI. EXPERIMENTS

To demonstrate the efficiency of the proposed block deficiency algorithm, we compared it with the quadtree algorithm proposed in [8] on randomly generated examples of hands and knowledge bases. As mentioned in the end of Sec. IV-A, the quadtree algorithm is incomplete when knowledge base is considered. To remedy this, we refined the quadtree algorithm by expanding the alphabet $\Sigma = \{1, 2, 3, 4\}$ to $\{1, 2, 3, 4, 5, 6\}$ and allow each node has up to six child nodes. We implemented both algorithms in Python3 and all experiments were executed on a laptop with Intel i7-6700 CPU and 16 GB RAM.

The colour set of a hand is the set of colours that tiles in the hand have. For example, the colour set of $H = (B1B1B1B3B5B6B8)(D1D2D3D5D5D5)$ is $\{B, D\}$. We say a hand is a pure hand if its colour set is a singleton. According to the number of different colours in the hand, we generate three experiment datasets, each contains 1000 random hands, and for each hand H , we randomly generate 30 or 100 knowledge bases such that they are compatible with H (i.e., no tile has five or more identical copies) and their sizes (i.e., number of tiles in the knowledge base) follow a normal distribution.

A. Pure hands

When examining pure hands, only tiles with the same colour in the knowledge base are concerned. As there are only 36 tiles with a given colour, they are at most 22 tiles in the knowledge base. Following a normal distribution, we generated 30 knowledge bases for each of the 1000 randomly generated pure hand. The experiments are summarised in Table I, where column 1 gives the range, say, [5,10), in which the number of tiles in KB is, column 2 shows the number of hand-KB pairs whose KB has size located in the range specified by column 1. Note we use ms (millisecond) as the base time unit. Columns 3-6 show the maximum or average time (in ms) of the quadtree and block deficiency algorithms. The last column shows the ratio of the average time of the quadtree algorithm to that of the block deficiency algorithm.

B. Hands with two colours

When examining hands with two colours, only tiles with one of the two colours in the knowledge base are concerned. As there are only 72 tiles with one of the two given colours, they are at most 58 tiles in the knowledge base. Following a normal

TABLE II
EXPERIMENTAL RESULTS OF 2-COLOUR HANDS

num. tiles in KB	num. hand/KB pairs	quadtree max. calc. time (ms)	block max. calc. time (ms)	quadtree avg. calc. time (ms)	block avg. calc. time (ms)	avg. ratio
[0, 10)	9963	424.5	9.4	15.1	0.5	28.3
[10, 20)	18798	1155.2	18.0	69.0	0.9	75.4
[20, 30)	25350	1633.0	18.1	156.1	1.1	145.6
[30, 40)	23646	2235.8	18.5	274.8	1.2	237.5
[40, 50)	15734	2525.1	17.4	381.8	1.2	311.7
[50, 58]	6509	2679.5	17.3	456.7	1.2	375.4
[0,58]	100,000	2679.5	17.3	208.8	1.0	200.4

TABLE III
EXPERIMENTAL RESULTS OF 3-COLOUR HANDS

num. tiles in KB	num. hand/KB pairs	quadtree max. calc. time (ms)	block max. calc. time (ms)	quadtree avg. calc. time (ms)	block avg. calc. time (ms)	avg. ratio
[0, 10)	4966	58.0	3.7	3.1	0.2	13.3
[10, 20)	8151	212.4	3.4	17.7	0.3	50.7
[20, 30)	11407	470.6	4.0	29.7	0.4	74.0
[30, 40)	14519	647.2	6.7	50.5	0.4	119.3
[40, 50)	15842	873.8	4.4	80.9	0.4	179.8
[50, 60)	15297	1054.8	5.4	110.8	0.5	242.1
[60, 70)	12611	1256.7	6.4	139.3	0.5	297.1
[70, 80)	9454	1325.8	4.4	165.8	0.5	349.6
[80, 94]	7753	1386.2	4.3	190.5	0.5	396.8
[0,94]	100,000	1386.2	6.7	90.1	0.4	209.6

distribution, we generated 100 knowledge bases for each of the 1000 randomly generated 2-colour hand. The experiments are summarised in Table II, where the meanings of the columns are the same as those in Table I.

C. Hands with three colours

When examining hands with three colours, all tiles in the knowledge base are concerned. As there are only 108 tiles, they are at most 94 tiles in the knowledge base. Following a normal distribution, we generated 100 knowledge bases for each of the 1000 randomly generated 3-colour hand. The experiments are summarised in Table III, where the meanings of the columns are the same as those in Table I.

D. Summary of experiment results

From our results we can see that the block deficiency algorithm is much more faster than the quadtree algorithm. In the worst cases, the quadtree algorithm takes 9.3 seconds to compute the deficiency, while the block deficiency algorithm takes at most 0.1 seconds. In average, the quadtree algorithm is 33x, 200x, and 210x slower than the block deficiency algorithm when the hands are pure, 2-colour, and 3-colour, respectively. While the quadtree algorithm is exact, it cannot be adopted by a Mahjong AI as in each step the Mahjong AI needs to respond within, say, 8 seconds while it often needs to call the quadtree algorithms dozens of times when

selecting a tile to discard. Our experiments also confirm that the block deficiency algorithm is exact when the deficiency of the hand w.r.t. a knowledge base is less than or equal to 4, which guarantees its effectiveness as a deficiency algorithm. Moreover, the results show that almost all ($\geq 99\%$) errors (i.e., the block deficiency is greater than the real deficiency) occur when the knowledge base contains less than or equal to 2 (out of 22), 10 (out of 58), and 20 (out of 94) tiles for pure hands, 2-colour hands, and 3-colour hands, respectively. In real games, this (error) rarely happens as when the knowledge base contains fewer tiles, the game is close to the end and the AI's hand should be close to ready (i.e., the deficiency should be close to 1).

We also tested the two deficiency algorithms in complete games. Assume that the four players use the same set of Mahjong AIs and they play 100 randomly generated games in two tests. In the first test, all Mahjong AIs use the quadtree algorithm. In average, each game was completed in 219.4 seconds; in the second test, all Mahjong AIs use the block algorithm. In average, each game was completed in 1.0 seconds.

VII. CONCLUSION

Computing how many tile changes are necessary to complete a Mahjong hand (i.e., the deficiency or shanten number) is a routine task for a Mahjong AI. There are very few published methods for this purpose and existing methods are either inefficient or not knowledge-aware. In this paper, we proposed an efficient algorithm for computing the deficiency of a Mahjong hand that respects the agent's knowledge of available tiles. The deficiency number calculated by this algorithm is always exact if the real deficiency is not larger than four. Experimental results on random hands and knowledge bases show that all errors occur when the real deficiency is larger than four and almost all ($\geq 99\%$) errors occur when the knowledge base contains very few tiles, which rarely occur simultaneously in practical games.

The new deficiency algorithm can be used as a basic procedure by any Mahjong AI in all variants of Mahjong. We have incorporated this algorithm in designing Mahjong AI for Sichuan Mahjong, which in average can complete a game in one second. Future work will exploit this advantage to generate huge game records and design and train strong Mahjong AI.

ACKNOWLEDGEMENT

This work was partially supported by National Science Foundation of China under Grant No. 12071271.

REFERENCES

- [1] Mahjong. Accessed: 2018-12-12.
- [2] Michael Bowling, Neil Burch, Michael Johanson, and Oskari Tammelin. Heads-up limit hold'em poker is solved. *Science*, 347(6218):145–149, 2015.
- [3] Noam Brown and Tuomas Sandholm. Safe and nested subgame solving for imperfect-information games. In Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems*, pages 689–699, 2017.
- [4] Shijing Gao and Shuqin Li. Bloody Mahjong playing strategy based on the integration of deep learning and XGBoost. *CAAI Transactions on Intelligence Technology*, n/a(n/a):1–12.
- [5] Qiqi Jiang, Kuangzheng Li, Boyao Du, Hao Chen, and Hai Fang. DeltaDou: Expert-level doudizhu AI through self-play. In Sarit Kraus, editor, *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019*, pages 1265–1271, 2019.
- [6] Moyuru Kurita and Kunihito Hoki. Method for constructing artificial intelligence player with abstractions to Markov Decision Processes in multiplayer game of Mahjong. *IEEE Transactions on Games*, 13(1):99–110, 2021.
- [7] Junjie Li, Sotetsu Koyamada, Qiwei Ye, Guoqing Liu, Chao Wang, Ruihan Yang, Li Zhao, Tao Qin, Tie-Yan Liu, and Hsiao-Wuen Hon. Suphx: Mastering Mahjong with deep reinforcement learning. *arXiv: 2003.13590*, 2020.
- [8] Sanjiang Li and Xueqing Yan. Let's Play Mahjong! *arXiv: 1903.03294*, 2019.
- [9] Naoki Mizukami and Yoshimasa Tsuruoka. Building a computer Mahjong player based on monte carlo simulation and opponent models. In *2015 IEEE Conference on Computational Intelligence and Games, CIG 2015*, pages 275–283. IEEE, 2015.
- [10] Arthur L. Samuel. Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3(3):210–229, 1959.
- [11] Claude E Shannon. Programming a computer for playing chess. In *Computer Chess Compendium*, pages 2–13. Springer, 1988.
- [12] David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Vedavyas Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy P. Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587):484–489, 2016.
- [13] Mingyan Wang, Tianwei Yan, Mingyuan Luo, and Wei Huang. A novel deep residual network-based incomplete information competition strategy for four-players Mahjong games. *Multim. Tools Appl.*, 78(16):23443–23467, 2019.
- [14] Kenshi Yoshimura, Teruhisa Hochin, and Hiroki Nomiya. Searching optimal movements in multi-player games with imperfect information. In *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*, pages 1–6, 2016.

APPENDIX

Proof of Lemma 2. (a) If $p = re = ke = 0$, π has no pair, and we cannot make a pair by using tiles in the remainder of π or the knowledge base. Thus π is incompletable.

(b) If $re = ke = 0$, then the only way we can have the eye is to use one pair in π to act as the eye. This is impossible if $p = 0$. In case $p > 0$, we need to create a new meld. However, by $rm = km = 0$, this is not possible either. Thus π is incompletable in this case.

(c) Suppose $m + n - e \leq 3$. In order to complete π , we need to create a new meld, but this is impossible if both rm and km are 0.

(d) Suppose $m + n \leq 3$, $p = ke = km = 0$ and $re = rm = 1$. If we cannot make the eye and a meld simultaneously, i.e., $em = 1$, π is still incompletable. \square

Proof of Lemma 3. By Lemma 1, we know $m + n > 4$ only if $p > 0$.

For case (a), we have $p > 0$ and $e \leq 1$. If $e = 1$, then all the other pairs in π are completable; if $e = 0$, we may use one pair in π as the eye. In either subcase, we need only complete $4 - m$ pmelds.

For case (b), if $e = 0$ and $re = 1$, the eye can be created by using a tile in the remainder and a tile from the knowledge base. Thus the cost is $4 - m + 1$. If $p > 0$ and $rm = 1$, we assign an existing pair as the eye and create a meld by starting with a tile in the remainder. The cost is also $4 - m + 1$ as we need to complete $4 - m - 1$ pmelds and create a meld from scratch.

In case (c), since neither condition of case (b) holds, we cannot complete the hand by making the eye or an meld by starting with a tile in the remainder. If $e = re = 0$ and $ke = 1$, we can make the eye by borrowing two tiles from KB ; if $p > 0$, $rm = 0$, and $km = 1$, we use one pair as the eye and make a new meld by borrowing three tiles from KB . In either case, the cost is $4 - m + 2$.

In case (d), none of the above holds. We have (i) $m + n = 4$, (ii) if $e = 0$ then $re = ke = 0$, and (iii) if $p > 0$ then $rm = km = 0$. There are three subcases. If $p > e = 0$, then $rm = km = re = ke = 0$. By Lemma 2 (b), π is incompletable. If $p = 0$, then $e = 0$ and hence $re = ke = 0$. By Lemma 2 (a), π is incompletable. If $p > 0$ and $e > 0$, then $rm = km = 0$ and $m + n - e \leq 3$. By Lemma 2 (c), π is incompletable. \square

Proof of Lemma 4. Except at most one pair, each pmeld in π can be completed by one tile change. To complete π , we need also to create at least one meld. In addition, if $p = 0$, we need also create the eye by using tiles in the remainder of π and KB . In case $e = 0$ but $p > 0$, we may assign an existing pair in π as the eye, but this incurs the cost of creating another meld from scratch.

Suppose we want to create a meld from scratch; that is, do not use any pmeld in π . We pick one tile from the remainder if $rm = 1$ and then complete it by borrowing two tiles in KB . When $rm = 0$, we need to create the meld by borrowing

three tiles from KB . This shows that $mcost$ is the minimum number of tile changes for creating such a meld.

The case of creating the eye is only slightly different and depends on the values of p, e, re and ke . If $re = 1$ or $ke = 1$, and we want to create the eye from scratch, then the cost is either 1 (if $re = 1$) or 2 (if $re = 0$ and $ke = 1$). That is, the cost to make the eye in this case is $ecost$.

In case (a), as $e = 1$, we already have the eye. The cost of π is $n - e$ (the number of completable pmelds in π) plus $mcost$, the minimal cost of making a meld from scratch.

In case (b), $p = e = 0$. We have $re = 1$ or $ke = 1$. Thus the cost of π is n (the number of completable pmelds in π) plus $mcost$ plus $ecost$.

In case (c), $p = 0$, $em = 1$, and $ke = 1$ or $km = 1$. From $em = 1$, we have $re = rm = 1$ and $ecost = 1$, $mcost = 2$, but we cannot make the eye and a new meld simultaneously from the remainder. By $ke = 1$ or $km = 1$, we have the cost is $n + mcost + ecost + 1 = n + 4$.

In case (d), $p > 0$ and $e = 0$. We may complete π by (i) creating a new meld and the eye from scratch, or (ii) assigning an existing pair in π as the eye and creating two new melds from scratch. We divide the discussion into subcases. As $p > 0$, it is possible that both re and ke are 0. Here we also note that $n \geq p > 0$.

- If $re = 1$, then $ecost = 1$ and the cost of π is $f_1 = n + mcost + 1$, which is at least 4 as $mcost \geq 2$. If the eye-meld conflict exists, the cost is at least 5.
- If $re = ke = 0$, then we can only make the eye by using an existing pair in π . Thus the cost of π is at least $f_2 = n - 1 + 2 \times mcost \geq 4$, as $mcost \geq 2$.
- Suppose $re = 0$, $ke = 1$ and $rm = 0$. Since $\max(rm, km) = 1$, we have $km = 1$ and, thus, $mcost = 3$, $ecost = 2$. Creating two melds from scratch requires at least 6 tile changes and thus it is cheaper to create the eye from scratch. In this case, the cost of π is at least $n + mcost + ecost = n + 5 \geq 6$.
- Suppose $re = 0$, $ke = 1$ and $rm = 1$. Then $mcost = 2$ and $ecost = 2$. If we choose to create the eye from scratch, then the cost is $n + mcost + ecost = n + 4$. Instead, suppose we assign one existing pair as the eye and make two melds from scratch. This incurs cost at least $n - 1 + 2 \times mcost = n + 3$. As a result, the cost of π in this case is at least $n + 3 \geq 4$.

It is then clear that the cost π in case (d) is at least the minimum of f_1 and f_2 and thus at least 4. \square