Prediction of Regular Search Tree Growth by Spectral Analysis

Stefan Edelkamp

Institut für Informatik
Georges-Köhler-Allee, Gebäude 51
79110 Freiburg
eMail: edelkamp@informatik.uni-freiburg.de

Abstract. The time complexity analysis of the IDA* algorithm has shown that predicting the growth of the search tree essentially relies on only two criteria: The number of nodes in the brute-force search tree for a given depth and the equilibrium distribution of the heuristic estimate. Since the latter can be approximated by random sampling, we accurately predict the number of nodes in the brute-force search tree for large depth in closed form by analyzing the spectrum of the problem graph or one of its factorization.
We further derive that the asymptotic brute-force branching factor is in fact the spectral radius of the problem graph and exemplify our considerations in the domain of the $(n^2 - 1)$-Puzzle.

1 Introduction

Heuristic search is essential to AI, since it allows very large problem spaces to be traversed with a considerably small number of node expansions. Nevertheless, storing this number of nodes in memory, as required in the A* algorithm [5], often exceeds the resources available. This is bypassed in an iterative deepening version of A*, IDA* for short, that searches the tree expansion of the original state graph instead of the graph itself. IDA* [10] applies bounded depth-first traversals with an increasing threshold on A*'s node evaluation function. The tree expansion may contain several duplicate nodes such that low memory consumption is counterbalanced with a considerably high overhead in time.

Fortunately, due to simple search tree pruning rules and expressive heuristic estimates to direct the search process, duplicates in regular search spaces are rare such that IDA* has been very successfully applied to solve solitaire games like the $(n^2 - 1)$ Puzzle [9, 12, 15] and Rubik's Cube [11].

Korf, Reid and Edelkamp [14] have analyzed the IDA* algorithm to predict the search performance of IDA* in the number of node expansions for a specific problem. The main result is that assuming consistency[1] of the integral heuristic

[1] Consistent heuristic estimates satisfy $h(v) - h(u) + 1 \geq 0$ for each edge (u, v) in the underlying problem graph. They yield monotone node evaluations $f(u) = g(u) + h(u)$ on generating paths with length $g(u)$. Admissible heuristics are lower bound estimates that underestimate the goal distance for each state. Consistent estimates are admissible.

estimate in the limit of large c, the expected total number of node expansions with cost threshold c in one iteration of IDA* is equal to

$$\sum_{d=0}^{c} n^{(d)} P(c-d),$$

where $n^{(d)}$ is the number of nodes in the brute-force search tree with depth d and P is the equilibrium distribution defined as the probability distribution of heuristic values in the limit of large depth. More precisely, $P(h)$ is the probability that a randomly and uniformly chosen node of a given depth has a heuristic value less than or equal to h. In practice the equilibrium distribution for admissible heuristic functions will be approximated by random sampling [13]; a representative sample of the problem space is drawn and classified according to the integral heuristic evaluation function. The value $n^{(d)}$ for large depths d without necessarily exploring the search tree, can be approximated with the asymptotic brute-force branching factor; the number of nodes at one depth divided by the number of nodes in the next shallower depth, in the limit as the depth goes to infinity. The asymptotic heuristic branching factor is defined analogously on search tree levels for two occurring values on the node evaluation function f. In some domains we observe anomalies in the limiting behavior of the asymptotic branching factors, e.g., in the (n^2-1)-Puzzle and odd values of n it alternates between two different values [2].

The observation that a consistent heuristic estimate h affects the relative depth to a goal instead of the branching itself is supported by the fact that IDA*'s exploration is equivalent to undirected iterative deepening exploration in a re-weighted problem graph with costs $1 + h(v) - h(u)$ for all edges (u, v). The new node evaluation $f'(u_j)$ of node u_j on path $p = (s = u_1, \ldots, u_t = t)$ equals $\sum_{i=1}^{j-1} (1 + h(u_{i+1}) - h(u_i))$ and telescopes to the old merit $f(u_j)$ minus $h(s)$. Therefore, the heuristic is best understood as a bonus to the search depth. Moreover, since we have only altered edge weights, it is not surprising that for bounded heuristic estimates and large depth the asymptotic heuristic branching factor equals the asymptotic brute-force branching factor.

Our main result in this paper is that in undirected problem graphs the value of the number of nodes in depth d of the brute-force search can be computed effectively by analyzing the spectrum of the adjacency matrix for the problem graph. The analysis requires some results of linear algebra and an algorithm of applied mathematics. Since the problem graph is considered to be large for regular search spaces we show how to factorize the problem graph through an equivalence relation of same branching behavior. We take the (n^2-1)-Puzzle as the running example, discuss the generality of the results from various points of view: other problem domains, general, especially undirected graph structures, and predecessor pruning. Finally, we give concluding remarks and shed light on future research options.

2 Linear Algebra Basics

Linear Mappings and Bases A mapping $f : V \to W$, with V, W being vector spaces over the field K (e.g. the set of real or the set of complex numbers) is *linear*, if $f(\lambda v + \mu w) = \lambda f(v) + \mu f(w)$ for all $v, w \in V$ and all $\lambda, \mu \in K$. A *basis* of a vector space V is a linear independent set of vectors that spans V. If the basis is finite, its cardinality defines the dimension $dim(V)$ of the vector space V, otherwise the dimension is said to be infinite.

Matrices and Basis-Transformations Linear mappings of vector spaces of finite dimension can be represented as matrices, since there is an isomorphism that maps the set of all $(m \times n)$ matrices to the set of all linear mappings from V to W according to their respectively fixed bases, where $dim(V) = n$ and $dim(W) = m$. Usually, V equals W and in this case the linear mapping f is called *endomorphism*. A *basis-transformation* from basis $\mathcal{A}$ to $\mathcal{B}$ in the vector space V can be represented by a transformation matrix $C_{\mathcal{AB}}$ which is the inverse of $C_{\mathcal{BA}}$. Very often, $\mathcal{A}$ is the canonical basis. Computing the inverse C^{-1} of a matrix C can be achieved by elementary row transformations, that convert the $(n \times 2n)$ matrix $[C \mid I]$ into $[I \mid C^{-1}]$, with I being the identity matrix.

Similarity and Normal Forms Two matrices A and B are *similar*, if there is a matrix C with $B = CAC^{-1}$. This is equivalent to the fact that there is an endomorphism f of V and two bases $\mathcal{A}$ and $\mathcal{B}$ with matrix A representing f according to $\mathcal{A}$ and B representing f according to $\mathcal{B}$. Similarity is an equivalence relation and one main problem in linear algebra is to derive a concise representative in the equivalence class of similar matrices, the *normal form*. A very simple form is the diagonal shape with non-zero values $\lambda_1, \ldots, \lambda_n$ only on the main diagonal. In this case, a matrix B is called *diagonizable* and can be written as $B = C \cdot \text{diag}(\lambda_1, \ldots, \lambda_n) \cdot C^{-1}$. Unfortunately, not all matrices are diagonizable, especially when the linear mapping is defined on the set of real numbers. Even if the vector space defining field is the set of complex numbers, only *tridiagonizability* can be granted, in which matrix A may have non-zero components above the main diagonal. Further simplifications lead to the so-called *Jordan normal form*.

Eigenvalues and Eigenspaces An endomorphism f of a vector space V over the field K contains an *eigenvalue* $\lambda \in K$, if there is a non-trivial vector $v \in V$, with $f(v) = \lambda v$. Any such non-trivial vector $v \in V$ with $f(v) = \lambda v$ is called *eigenvector*. If there is a basis $\mathcal{B}$ of eigenvectors, then the matrix representation according to $\mathcal{B}$ has a diagonal shape. In this case f is also called *diagonizable*. It can be shown that the eigenvalues are roots of the *characteristic equation* $P_f(\lambda) = \det(A - \lambda I) = 0$, where the determinant det(A) is defined as $\sum_{\sigma \in S_n} \prod_{i<j} (\sigma(j) - \sigma(i))/(j-i)) \cdot a_{1\sigma(1)} \cdot \ldots \cdot a_{n\sigma(n)}$ with S_n being the set of all n-permutations. If the polynomial $P_f(\lambda)$ factorizes, i.e. $P_f(\lambda) = const \cdot \prod_{i=1}^{k} (\lambda - \lambda_i)$, which is the case for matrices of complex numbers, the corresponding eigenspaces $E_f(\lambda_i)$ have to be computed. If then the number of occurring linear terms $(\lambda - \lambda_i)$

in $P_f(\lambda)$, the *algebraic multiplicity* of λ_i, equals the dimension of $E_f(\lambda_i)$, the *geometric multiplicity* of λ_i, then A is indeed *diagonizable*.

3 Partitioning the Search Space

The (n^2-1)-Puzzle is a sliding tile toy problem. It consists of (n^2-1) numbered tiles that can be slid into a single empty position, called the blank. The goal is to rearrange the tiles such that a certain goal position is reached. Figure 1 depicts possible end configurations of well-known instances to the (n^2-1)-Puzzle: For $n = 3$ we get the Eight-Puzzle, for $n = 4$ the Fifteen-Puzzle and for $n = 5$, the Twenty-Four-Puzzle is met. The state spaces for these problems grow

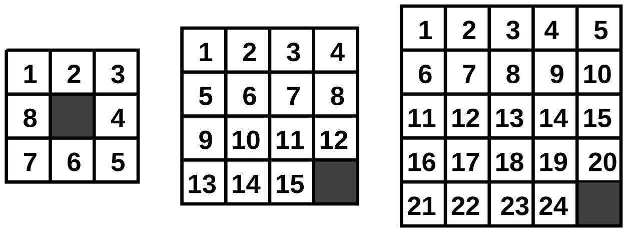

Fig. 1. The Eight-, Fifteen- and Twenty-Four-Puzzles.

exponentially. The exact number of reachable states (independent of the initial one) is $(n^2)!/2$ which resolves to approximately 10^5 states for the Eight-Puzzle, 10^{13} states in the Fifteen-Puzzle and 10^{25} states in the Twenty-Four-Puzzle.

We partition the search space S in classes $S_1, \ldots, S_k$, collecting states into groups with same branching behavior. In other words we devise an equivalence relation that partitions the state space into equivalence classes: two states are equivalent if their long term branching behavior coincides. All states in one equivalence class S_i, $i \in \{1, \ldots, k\}$, necessarily have the same node branching factor, defined as the number of children a node has in the brute-force search tree.

For the example of the (n^2-1)-Puzzle a partition is given by the following relation: two states are equivalent if the blank is at the same absolute position. Obviously the subtrees of such nodes are isomorphic, since the branching behavior of equivalent states has to be the same. A further reduction of the search tree is established by partitioning the search space with respect to symmetry. For the (n^2-1) Puzzle we establish three branching types: corner or c-nodes with node branching factor 2, side or s-nodes with node branching factor 3, and middle or m-nodes with node branching factor 4. However, does the long time node branching behavior depend on these node types only? In the Eight- and Fifteen-Puzzles this is the case, since for symmetry reasons all c, s and m nodes generate the same subtree structure. For the Twenty-Four-Puzzle, however, the

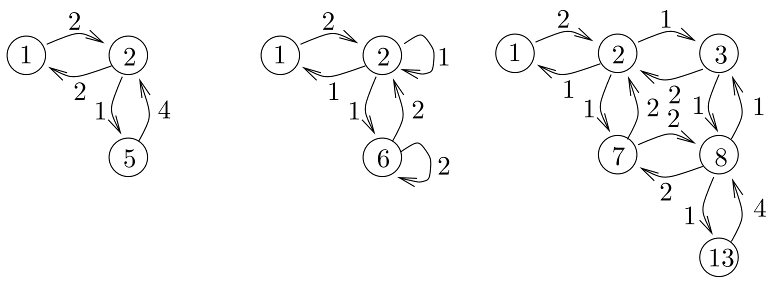

Fig. 2. Equivalence Graph for the Eight-, Fifteen- and Twenty-Four-Puzzles.

search tree of two side or two middle states may differ. For this case we need six classes with a blank at position 1,2,3,7,8, and 13 according to the tile labeling in Figure 2. In the general case the number of different node branching classes in the (n^2-1) Puzzle is

$$\sum_{i=0}^{\lceil n/2 \rceil} i = \binom{\lceil n/2 \rceil}{2} = \lceil n/2 \rceil (\lceil n/2 \rceil - 1)/2.$$

This still compares well to a partition according to the n^2 equivalent classes in the first factorization (savings of a factor of about eight) and of course to the $(n^2)!/2$ states in the overall search space (exponential savings).

4 Equivalence Graph Structure

Utilizing this partition technique we define the weighted *equivalence graph* $\overline{G} = (\overline{V}, \overline{E}, w)$ as follows. The set of nodes $\overline{V}$ equals the set of equivalence classes and an edge e from class $S_i \in \overline{V}$ to $S_j \in \overline{V}$ with weight $w(e)$ is drawn, if every state in S_i leads to w states in class S_j. Obviously, the sum of all outgoing edges equals the node branching factor. Let $A_{\overline{G}}$ be the adjacency matrix with respect to the *equivalence graph* $\overline{G}$. Since the explorations in G and $\overline{G}$ span the same search-tree structure the search tree growth will be the same.

A generator matrix P for the population of nodes according to the given equivalence relation is defined by $P = A_{\overline{G}}^T$. More precisely, $P_{j,i} = l$ if a node of type i in a given level leads to l nodes of type j in the next level. We immediately infer that $N^{(d)} = PN^{(d-1)}$, with $N^{(d)}$ being the vector of nodes in depth d of the search tree. If $|| \cdot ||_1$ denotes the vector norm $||x||_1 = |x_1| + \ldots + |x_k|$ then the number of nodes $n^{(d)}$ in depth d is equal to $||N^{(d)}||_1$.

The asymptotic branching factor b (if it exists) is defined as the limit of $n^{(d)}/n^{(d-1)}$ for increasing d and equals the weighted product of the node frequencies $b = \sum_{i=1}^{k} b_i f_i$, where f_i is the fraction of nodes of class i with respect to the total number of nodes. As we will see, we can compute the branching factor analytically without actually determining node frequency values.

The first observation is that in case of convergence the asymptotic branching factor is not only met in the the overall search tree expansion but in every equivalence class. Since all frequencies of nodes converge we have that $b = \lim_{d\to\infty} N_i^{(d)}/N_i^{(d-1)}$, with $N_i^{(d)}$ being the number of nodes of class i in depth d, $i \in \{1,\ldots,k\}$. In other words, if the ratio of the cardinality of one equivalence class and the overall search space size settles and the search space size grows with factor b, then the equivalence class size itself grows with factor b.

We represent the fractions f_i as a distribution vector F. We first assume that this vector converges in the limit of large depth. The considerations for an analytical solution to the branching factor problem result in the equations $bF = FP$, where b is the asymptotic branching factor. In addition, we have the equation that the total of all node frequencies is one. The underlying mathematical issue is an eigenvalue problem. Transforming $bF = PF$ leads to $0 = (P - bI)F$ for the identity matrix I. The solutions for b are the roots of the characteristic equation $\det(P - bI) = 0$ where det is the determinant of the matrix. Since $\det(P - bI) = \det(P^T - bI)$ the transposition of the equivalence graph matrix $A_{\overline{G}}$ preserves the value of b. In case of the Eight-Puzzle $\det(P - bI)$ equals

$$\det\begin{pmatrix} 0-b & 2 & 0 \\ 2 & 0-b & 1 \\ 0 & 4 & 0-b \end{pmatrix} = 0.$$

This equation is equivalent to $b(4-b^2)+4b = 0$, yielding the following three solutions $-\sqrt{8} = -2.828427124$, 0, $\sqrt{8} = 2.828427124$. Experimental results show that the branching factor alternates every two depth values between 3 and $8/3 = 2.666666666$. Since $\sqrt{8}$ is the geometric mean of 3 and $8/3$ the value $\sqrt{8}$ is the proper choice for the asymptotic branching factor b of the brute-force search tree.

For the case of the Fifteen-Puzzle we have to calculate

$$\det\begin{pmatrix} 0-b & 2 & 0 \\ 1 & 1-b & 1 \\ 0 & 2 & 2-b \end{pmatrix} = 0,$$

which simplifies to $(1-b)(b-2)b + 4b - 4 = 0$. The solution to this equation are 1, $1+\sqrt{5} = 3.236067978$, and $1-\sqrt{5} = -1.236067978$. The value $1+\sqrt{5}$ matches experimental data for the asymptotic branching factor.

For the Twenty-Four-Puzzle we have to solve

$$\det\begin{pmatrix} 0-b & 2 & 0 & 0 & 0 & 0 \\ 1 & 0-b & 1 & 1 & 0 & 0 \\ 0 & 2 & 0-b & 0 & 1 & 0 \\ 0 & 2 & 0 & 0-b & 2 & 0 \\ 0 & 0 & 1 & 2 & 0-b & 1 \\ 0 & 0 & 0 & 0 & 4 & 0-b \end{pmatrix} = 0.$$

The six eigenvalues are 0, 0, $\sqrt{3} = 1.732050808$, $-\sqrt{3} = -1.732050808$, $\sqrt{12} = 3.464101616$, and $-\sqrt{12} = -3.464101616$. Experiments show that for large depth the branching factor oscillates and that the geometric mean is 3.464101616.

We conclude that the asymptotic branching factor in the example problems is the *largest* eigenvalue of the adjacency matrix for the equivalence graph and that we observe anomalies if the largest eigenvalue has a negative counterpart of the same absolute value. In the following we will analyze the structure of the eigenvalue problem to show why this is the case.

5 Exact Prediction of Search Tree Size

The equation $N^{(d)} = PN^{(d-1)}$ can be unrolled to $N^{(d)} = P^d N^{(0)}$. We briefly sketch how to compute P^d for large d. We have seen that P is *diagonizable*, if there exists a invertible matrix C and a diagonal matrix Q with $P = CQC^{-1}$. This simplifies the calculation of P^d, since we have $P^d = CQ^dC^{-1}$ (the remaining terms $C^{-1}C$ cancel). By the diagonal shape of Q the value of Q^d is obtained by taking the matrix elements $q_{i,i}$ to the power of d. These elements are the eigenvalues of P. This connection is not surprising, since in case of convergence of the vector of node frequencies F we have seen that the branching factor itself is a solution to the eigenvalue problem $PF = bF$. We conclude that in case of *diagonizability* we can exactly predict the number of nodes of depth d by determining the set of eigenvalues of P.

In the example of the Eight-Puzzle the eigenvectors for the eigenvalues $-\sqrt{8}$, 0, and $\sqrt{8}$ are $(2, -\sqrt{8}, 1)^T$, $(-2, 0, 1)^T$, and $(2, \sqrt{8}, 1)^T$, respectively. Therefore, the basis-transformation matrix C is given by

$$C = \begin{pmatrix} 2 & -2 & 2 \\ -\sqrt{8} & 0 & \sqrt{8} \\ 1 & 1 & 1 \end{pmatrix}$$

with the following inverse

$$C^{-1} = 1/16 \begin{pmatrix} 2 & -\sqrt{8} & 4 \\ -2 & 0 & 8 \\ 2 & \sqrt{8} & 4 \end{pmatrix}.$$

With $Q = diag(-\sqrt{8}, 0, \sqrt{8})$ we have $C^{-1}C = I$ and $C^{-1}PC = Q$. Therefore, calculating $N^{(d)} = P^d N^{(0)}$ for $d \geq 1$ corresponds to $N^{(d)} = CQ^dC^{-1}N^{(0)}$, where $Q^d = \text{diag}((-\sqrt{8})^d, 0, (\sqrt{8})^d)$. Hence, $N^{(d)}$ equals to

$$1/16 \begin{pmatrix} 2 & -2 & 2 \\ -\sqrt{8} & 0 & \sqrt{8} \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} (-\sqrt{8})^d & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & (\sqrt{8})^d \end{pmatrix} \begin{pmatrix} 2 & -\sqrt{8} & 4 \\ -2 & 0 & 8 \\ 2 & \sqrt{8} & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

which resolves to

$$N^{(d)} = 1/16 \left(4\sqrt{8}^d((-1)^d + 1), 2\sqrt{8}^{d+1}((-1)^{d+1} + 1), 2\sqrt{8}^d((-1)^d + 1)\right)^T.$$

The exact formula for $N^{(d)}$ and small values of d validates the observed search tree growth: $N^{(1)} = (0,2,0)^T$, $N^{(2)} = (4,0,2)^T$, $N^{(3)} = (0,16,0)^T$, $N^{(4)} = (32,0,16)^T$, etc.

The closed form for $N^{(d)}$ explicitly states that the asymptotic branching factor for the Eight Puzzle is $\sqrt{8}$. Moreover, the odd-even effect for branching in that puzzle is established by the factor $(-1)^d + 1$, which cancels for an odd value of d. Nevertheless, solving the characteristic equation and establishing the basis of eigenvectors by hand is tedious work. Fortunately, the application of symbolic mathematical tools such as Maple and Mathematica help to perform the calculations in larger systems.

For the Fifteen-Puzzle the basis-transformation matrix C and its inverse C^{-1} are

$$C = \begin{pmatrix} 1 & -1 & 1 \\ 1-\sqrt{5} & -1 & 1+\sqrt{5} \\ 3/2-1/2\sqrt{5} & 1 & 3/2+1/2\sqrt{5} \end{pmatrix}$$

and

$$C^{-1} = \begin{pmatrix} 1/50\,(5+3\sqrt{5})\sqrt{5} & -1/50\,(5+\sqrt{5})\sqrt{5} & 1/5 \\ -2/5 & -1/5 & 2/5 \\ 1/50\,(-5+3\sqrt{5})\sqrt{5} & -1/50\,(-5+\sqrt{5})\sqrt{5} & 1/5 \end{pmatrix}.$$

The vector of node counts is

$$N^{(d)} = \begin{pmatrix} 1/50\,(1-\sqrt{5})^d(5+3\sqrt{5})\sqrt{5}+2/5+ \\ 1/50\,(1+\sqrt{5})^d(-5+3\sqrt{5})\sqrt{5} \\[1ex] 1/50\,(1-\sqrt{5})(1-\sqrt{5})^d(5+3\sqrt{5})\sqrt{5}+2/5+ \\ 1/50\,(1+\sqrt{5})(1+\sqrt{5})^d(-5+3\sqrt{5})\sqrt{5} \\[1ex] 1/50\,(3/2-1/2\sqrt{5})(1-\sqrt{5})^d(5+3\sqrt{5})\sqrt{5}-2/5+ \\ 1/50\,(3/2+1/2\sqrt{5})(1+\sqrt{5})^d(-5+3\sqrt{5})\sqrt{5} \end{pmatrix}$$

such that the exact total number of nodes in depth d is

$$1/50\left(7/2-3/2\sqrt{5}\right)\left(1-\sqrt{5}\right)^d\left(5+3\sqrt{5}\right)\sqrt{5}+2/5+$$

$$1/50\left(7/2+3/2\sqrt{5}\right)\left(1+\sqrt{5}\right)^d\left(-5+3\sqrt{5}\right)\sqrt{5}$$

The number of corner nodes (1,0,2,2,10,26,90,...), the number of side nodes (0,2,2,10,26,90,282,...) and the number of middle nodes (0,0,6,22,70,230,...) grow as expected. The largest eigenvalue $1+\sqrt{5}$ dominates the growth of the search tree in the limit for large d.

In the Twenty-Four-Puzzle the value $N^{(d)}$ equals

$$\begin{pmatrix} 1/36\,\left(-2\sqrt{3}\right)^d + 2/9\,\left(-\sqrt{3}\right)^d + 2/9\,\left(\sqrt{3}\right)^d + 1/36\,\left(2\sqrt{3}\right)^d \\ -1/18\sqrt{3}\left(-2\sqrt{3}\right)^d - 2/9\sqrt{3}\left(-\sqrt{3}\right)^d + 2/9\sqrt{3}\left(\sqrt{3}\right)^d + 1/18\sqrt{3}\left(2\sqrt{3}\right)^d \\ 1/18\,\left(-2\sqrt{3}\right)^d + 1/9\,\left(-\sqrt{3}\right)^d + 1/9\,\left(\sqrt{3}\right)^d + 1/18\,\left(2\sqrt{3}\right)^d \\ 1/12\,\left(-2\sqrt{3}\right)^d + 1/12\,\left(2\sqrt{3}\right)^d \\ -1/18\sqrt{3}\left(-2\sqrt{3}\right)^d + 1/9\sqrt{3}\left(-\sqrt{3}\right)^d - 1/9\sqrt{3}\left(\sqrt{3}\right)^d + 1/18\sqrt{3}\left(2\sqrt{3}\right)^d \\ 1/36\,\left(-2\sqrt{3}\right)^d - 1/9\,\left(-\sqrt{3}\right)^d - 1/9\,\left(\sqrt{3}\right)^d + 1/36\,\left(2\sqrt{3}\right)^d \end{pmatrix}$$

for the following total of nodes in depth d

$$n^{(d)} = 1/36\,\left(7-4\sqrt{3}\right)\left(-2\sqrt{3}\right)^d + 1/9\,\left(2-\sqrt{3}\right)\left(-\sqrt{3}\right)^d +$$

$$1/9\,\left(2+\sqrt{3}\right)\left(\sqrt{3}\right)^d + 1/36\,\left(7+4\sqrt{3}\right)\left(2\sqrt{3}\right)^d.$$

The value for small d validates that the total number of nodes increases as expected (2,6,18,60,198,684,...). Once again the vector of the largest absolute value determines the search tree growth.

If the size of the system is large, the exact value of $N^{(d)}$ has to be approximated. One option to bypass the intense calculations for determinants of large matrices and roots of high-degree polynomials is to compute the asymptotic branching factor b. The number of nodes in the brute-force search tree is then approximated by $n^{(d)} \approx b^d$.

6 Approximate Prediction of Search Tree Size

The matrix denotation for calculating the population of nodes according to the given equivalence relation implies $N^{(d)} = PN^{(d-1)}$, with $N^{(d)}$ being the vector of equivalent class sizes. The asymptotic branching factor b is given by the limit of $||N^{(d)}||_1/||N^{(d-1)}||_1$ which equals $N_i^{(d)}/N_i^{(d-1)}$ in any component $i \in \{1,\ldots,k\}$. Evaluating $N_i^{(d)}/N_i^{(d-1)}$ for increasing depth d is exactly what is considered in the algorithm of van Mises for approximating the largest eigenvalue (in absolute terms) of P. The algorithm is also referred to as the *power iteration* method.

As a precondition, the algorithm requires that P be diagonizable. This implies that we have n different eigenvalues $\lambda_1,\ldots,\lambda_n$ and each eigenvalue λ_i with multiplicity of α_i has α_i linear independent eigenvectors. Without loss of generality, we assume that the eigenvalues are given in decreasing order $|\lambda_1| \geq |\lambda_2| \geq \ldots \geq |\lambda_k|$. The algorithm further requires that the start vector $N^{(0)}$ have a representation in the basis of eigenvectors in which no coefficient according to λ_1 is trivial.

We distinguish the following two cases: $|\lambda_1| > |\lambda_2| \geq \ldots \geq |\lambda_k|$ and $|\lambda_1| = |\lambda_2| > \ldots \geq |\lambda_k|$. In the first case we obtain that (independent of the choice of

$j \in \{1, \ldots, k\}$) the value of $\lim_{d\to\infty} N_j^{(d)}/N_j^{(d-1)}$ equals $|\lambda_1|$. Similarly, in the second case $\lim_{d\to\infty} N_j^{(d)}/N_j^{(d-2)}$ is in fact λ_1^2. The cases $|\lambda_1| = \ldots = |\lambda_l| > \ldots \geq |\lambda_k|$ for $l > 2$ are dealt with analogously. The outcome of the algorithm and therefore the limit in the number of nodes in layers with difference l is $|\lambda_1|^l$, so that once more the geometric mean turns out to be $|\lambda_1|$.

We indicate the proof of the first case only. Diagonizability implies a basis of eigenvectors $b_1 \ldots, b_k$. Due to $|\lambda_1| > |\lambda_2| \geq \ldots \geq |\lambda_n|$ the quotient of $|\lambda_i/\lambda_1|^d$ converges to zero for large values of d. If the initial vector $N^{(0)}$ with respect to the eigenbasis is given as $x_1b_1 + x_2b_2 + \ldots + x_kb_k$ applying P^d yields $x_1P^db_1 + x_2P^db_2 + \ldots + x_kP^db_k$ by linearity of P, which further reduces to $x_1b_1\lambda_1^d + \lambda_2^dx_2b_2 + \ldots + \lambda_n^dx_kb_k$ by the definition of eigenvalues and eigenvectors. The term $x_1b_1\lambda_1^d$ will dominate the sum for increasing values of d. Factorizing λ_1^d in the numerator and λ_1^{d-1} in denominator of the quotient of $N_j^{(d)}/N_j^{(d-1)}$ results in an equation of the form $x_1b_1\lambda_1 + R$ where $\lim_{d\to\infty} R$ is bounded by a constant, since except of the leading term $x_1b_1\lambda_1$ both numerator and denominator in R only involve expressions of the form $O(|\lambda_i/\lambda_1|^d)$. Therefore, to find the asymptotic branching factor analytically, it suffices to determine the set of eigenvalues of P and to take the largest one. This corresponds to the results of the asymptotic branching factors in the $(n^2 - 1)$-Puzzles.

In the Eight-Puzzle the ratio $N_1^{(d)}/N_1^{(d-2)}$ is equal to 8 for $d > 3$ and, therefore, approximates λ_1^2. The value $n^{(d)}/\sqrt{8}^d$ alternates between $3/4$ and $1/\sqrt{2}$. Hence, $\sqrt{8}^d$ approximates the search tree growth.

For the Fifteen-Puzzle for increasing depth d the value $N_1^{(d)}/N_1^{(d-1)}$ equals 1, 3, 13/5, 45/13, 47/15, 461/141, 1485/461, 4813/1485, 15565/4813, 50381/15565, 163021/50381, 527565/163021 = 3.236178161, etc., a sequence approximating $1+\sqrt{5} = 3.236067978$. Moreover, the ratio of $n^{(d)}$ and $(1+\sqrt{5})^d$ quickly converges to $1/50\,\left(7/2 + 3/2\,\sqrt{5}\right)\left(-5 + 3\,\sqrt{5}\right)\sqrt{5} = .5236067984$.

In the Twenty-Four-Puzzle the ratio $N_1^{(d)}/N_1^{(d-2)}$ converges to 12 starting with the sequence 6, 9, 11, 129/11, 513/43, 683/57, 8193/683, 32769/2731, 43691/3641 = 11.99972535, etc. The quotient $n^{(d)}/\sqrt{12}^d$ for larger depth alternates between .3888888889 and .3849001795 and is therefore bounded by a small constant.

If n is even – as in the Fifteen-Puzzle – the largest eigenvalue is unique and if n is odd – as in the Eight- and in the Twenty-Four-Puzzle – we find two eigenvalues with the same absolute value verifying that every two depths the node sizes will asymptotically increase by the square of these values.

7 Generalizing the Approach

Iterating the algorithm with $||N^{(d)}||_1/||N^{(d-1)}||_1$ instead of $N_j^{(d)}/N_j^{(d-1)}$ shows that the convergence conditions according to G and $\overline{G}$ are equivalent. This is important, since other graph properties may alter, e.g. symmetry of A_G is not inherited by $A_{\overline{G}}$. Therefore, we concentrate on diagonizability results of A_G, which are easier to obtain. The *Theorem of Schur* states that symmetric matrices are indeed diagonizable. Moreover, the eigenvalues are real and the matrix to perform the basis transformation has the eigenvectors in its columns.

For the (n^2-1)-Puzzle we are done. Since G is undirected, A_G is indeed symmetric. In the spectrum of A_G power iteration either obtains a unique branching factor $b = |\lambda_1|$ or a branching factor of λ_1^2 for every two iterations. Therefore, the branching factor is the *spectral radius* $\rho = |\lambda_1|$.

7.1 Other Problem Spaces

Since the search tree is often exponentially larger than the problem graph we have reduced the prediction of the search tree growth to the spectral analysis of the explicit adjacency representation of the graph. As long as this graph is available, accurate and approximate predictions for the brute-force and subsequently for the heuristic search tree growth can be computed.

However, the calculations for large implicitly given graphs are involved such that reduction of the analysis to a smaller structure is desirable. For the (n^2-1)-Puzzle we proposed a compression to a few branching classes. The application of equivalence class reduction to exactly predict the search tree growth relies on the regular structure of the problem space. This technique is available as long as the same branching behavior for different states is given.

For *Rubik's Cube* without predecessor elimination $N^{(d)}$ equals 18^d since all nodes in the search tree span a complete 18-ary subtree. With predecessor elimination the node branching factor reduces to 15, since for each of the three twists *single clockwise*, *double clockwise*, and *counterclockwise* there is a remainder of five sides *front*, *back*, *right*, *left*, *up*, and *down* that are available. If we further restrict rotation of opposite sides to exactly one order we get the transition matrix $((6\ 6), (9\ 6))$, where the first class is the set of primary nodes with branching factor 15, and the second class is the class of secondary nodes with branching factor 12. The eigenvalues are $6 + 3\sqrt{6}$ and $6 - 3\sqrt{6}$ and the value $n^{(d)}$ equals $1/2\ \left(6 + 3\,\sqrt{6}\right)^d + 1/2\ \left(6 - 3\,\sqrt{6}\right)^d$. For small values of d experimental data as given in [11] matches this analytical study. The observed asymptotic branching factor is $6 + 3\sqrt{6} = 13.34846923$ as expected.

Extending the work to problem domains like the PSPACE-complete *Sokoban* problem [1] is challenging. It is difficult to derive an accurate prediction, since the branching behavior of the tree includes almost all state facets. Therefore, a more complicated search model has to be devised to derive exact or approximate search tree prediction in this domain. As Andreas Junghanns has coined in his Ph.D. dissertation [8], the impact of the search tree node prediction formula $\sum_{d=0}^{c} n^{(d)} P(c-d)$ has still to be shown. In the other PSPACE-complete sliding block game *Atomix* [7, 6] simplification based on branching equivalences do apply and yield savings that are exponential in the number of atoms, but this void labeling scheme still results in an intractable size of the equivalence graph structure. Only very small games can be analyzed by this method.

7.2 Pruning

When incorporating pruning to the exploration process, symmetry of the underlying graph structure may be affected. Once again we consider the Eight-Puzzle. The adjacency matrix $A_{\overline{G}}^{pred}$ for predecessor elimination now consists of four classes: *cs*, *sc*, *mc* and *cm*, where the class *ij* indicates that the predecessor of a j-node in the search tree is an i node.

$$A_{\overline{G}}^{pred} = \begin{pmatrix} 0\,1\,0\,0 \\ 1\,0\,0\,1 \\ 2\,0\,0\,0 \\ 0\,0\,3\,0 \end{pmatrix}$$

In this case we cannot infer diagonizability according to the set of real numbers. Fortunately, we know that the branching factor is a positive real value since the iteration process is real. Therefore, we may perform all calculation to predict the search tree growth with complex numbers, for which the characteristic polynomial factorizes. The branching factor and the search tree growth can be calculated analytically and the iteration process eventually converges.

In the example, the set of (complex) eigenvalues is $i\sqrt{2}$, $-i\sqrt{2}$, $\sqrt{3}$, and $-\sqrt{3}$. Therefore, the asymptotic branching factor is $\sqrt{3}$. The vector $N^{(d)}$ is equal to

$$\begin{pmatrix} 1/5\,\left(i\sqrt{2}\right)^d + 1/5\,\left(-i\sqrt{2}\right)^d + 3/10\,\left(\sqrt{3}\right)^d + 3/10\,\left(-\sqrt{3}\right)^d \\ -1/10\,i\sqrt{2}\left(i\sqrt{2}\right)^d + 1/10\,i\sqrt{2}\left(-i\sqrt{2}\right)^d + 1/10\sqrt{3}\left(\sqrt{3}\right)^d - 1/10\sqrt{3}\left(-\sqrt{3}\right)^d \\ 3/20\,i\sqrt{2}\left(i\sqrt{2}\right)^d - 3/20\,i\sqrt{2}\left(-i\sqrt{2}\right)^d + 1/10\sqrt{3}\left(\sqrt{3}\right)^d - 1/10\sqrt{3}\left(-\sqrt{3}\right)^d \\ -1/10\,\left(i\sqrt{2}\right)^d - 1/10\,\left(-i\sqrt{2}\right)^d + 1/10\,\left(\sqrt{3}\right)^d + 1/10\,\left(-\sqrt{3}\right)^d \end{pmatrix}.$$

Finally, the total number of nodes in depth d is

$$n^{(d)} = 1/5\,\left(1/2 + 1/4\,i\sqrt{2}\right)\left(i\sqrt{2}\right)^d + 1/5\,\left(1/2 - 1/4\,i\sqrt{2}\right)\left(-i\sqrt{2}\right)^d +$$
$$1/10\,\left(4 + 2\sqrt{3}\right)\left(\sqrt{3}\right)^d + 1/10\,\left(4 - 2\sqrt{3}\right)\left(-\sqrt{3}\right)^d.$$

For small values of d the value $n^{(d)}$ equals 1, 2, 4, 8, 10, 20, 34, 68, 94, 188 etc.

7.3 Non-Diagonizability

Since we assumed diagonizability, the eigenspaces $L(\lambda_i)$ according to the values λ_i have full rank α_i. In general this is not true. Not all matrices are diagonalizable. In this case the best thing one can do is to transform the matrix into *Jordan Form* which has blocks on the diagonal, each block being $r \times r$, with the eigenvalue on the diagonal, 1's above the diagonal and 0's everywhere else.

More precisely, a matrix A has *Jordan Form* J for an invertible matrix T, if $J = T^{-1}AT$ consists of so-called Jordan-blocks $J_1, \ldots, J_m$. One Jordan-block has an eigenvalue on the main diagonal and 1s on the diagonal above. Therefore, T gives a basis of eigenvectors and so-called *main vectors.* Each Jordan-block J_l of dimension j_l corresponds to one eigenvector t_1 and $j_l - 1$ main vectors $t_2, \ldots, t_{j_l}$ with $(A - \lambda_i I)t_0 = 0$ and $(A - \lambda_i I)t_m = t_{m-1}$, $m = 2, \ldots, j_l$. Using the Jordan basis one can devise $P^d N^{(0)}$ similar to the case above.

7.4 Start Vector

The second subtlety arises even if the matrix is diagonalizable. We are interested in determining the behavior of $P^d N^{(0)}$ for large d, where P is an $n \times n$ matrix and $N^{(0)}$ is an $n \times 1$ vector. Suppose that P is diagonalizable, which means that there is a basis of eigenvectors. Hence, $N^{(0)}$ can be written as a sum of eigenvectors: $N^{(0)} = v_1 + v_2 + v_3 + \ldots + v_n$ where v_i is an eigenvector with eigenvalue λ_i. It follows that $P^d N^{(0)} = \lambda_1^d v_1 + \lambda_2^d v_2 + \lambda_3^d v_3 + ... + \lambda_n^d v_n$ So the term with the largest corresponding eigenvalue will dominate for large d, *provided that the eigenvector is non-zero.* It may happen that the initial vector v has component of zero in the eigenspace of the largest eigenvalue. In general, the algorithm finds the largest eigenvalue in which the corresponding component is non-zero.

Fortunately, this observation is more theoretical in nature. In the iteration process this case is very rarely fulfilled. Rounding errors will soon or later lead to non-zero components. Moreover, to determine the asymptotic branching factor we have several initial states to choose from such that at least one has to yield non-zero coefficients.

8 Previous Work

This paper extends the work of Edelkamp and Korf [2] that already derived the asymptotic branching factors of the sliding-tile puzzles and Rubik's Cube. However, their approach lack sufficient convergence conditions. We established the criterion of diagonizability of the adjacency graph matrix of the problem graph that emerges of the algorithm of van Mises and showed that this criterion is fulfilled in undirected graphs by the Theorem of Schur. The $(n^2 - 1)$-Puzzles and Rubik's Cube are chosen to illustrate the techniques, since they are inherently difficult to solve and often considered in case studies.

The set of recurrence relations in [2] also showed that the numbers of nodes at various depths can be calculated in time linear to the product of the number of node classes and the depth of search tree by numerically iterating the recurrence relations. In contrast to this finding, the current paper resolves the problem of how to compute a closed form for the number of nodes. Last but not least, the given mathematical formalization of equivalence classification, diagonalization and power iteration builds a bridge for more powerful results in applying known mathematical theorems. At least in theory, generality to different problem spaces is given, since this approach applies to any problem graph with a diagonizable matrix and probably to more than that.

9 Conclusion and Discussion

In the paper we have improved the prediction of the number of node expansions in IDA* by an exact derivation of the number of nodes in the brute-force search tree. We have resolved the question of convergence to explain anomalies in of the asymptotic branching factor. The asymptotic branching factor is the spectral radius of the successor generation matrix and can be computed with the power iteration method. The approach extends to further regular problem spaces and can cope with simple pruning rules. The main result is that diagonizability is granted in undirected problem graphs, such that exact and approximate calculation of the brute-force search-tree are mathematically sound. The technique for establishing a closed form is not standard, and it is hard to suggest other methodologies to actually solve the set of recurrence relations.

Moreover, given the adjacency matrix P of an undirected graph by analyzing $N_i^{(d)}/N_i^{(d-1)}$ and $N_i^{(d)}/N_i^{(d-k)}$, $k > 1$, of the equation $N^{(d)} = N^{(d-1)}P$ gives the (mean) asymptotic branching factor. This is in fact the algorithm of van Mises to determine the largest eigenvalue of P for whose applicability we have to test if P is diagonizable. The paper closes the small gap in literature to accurately predict search tree growth in closed form and to compute the branching factor both analytically and numerically without relying on strong experimental assumption on the convergence.

Since for practical problems in which IDA* applies it is very unlikely that the entire graph structure can be kept in main memory, the approach helps only if some reduction of the branching behavior with respect to equivalence classes can be obtained. Therefore, the analysis is limited to the cases where the the successor generator matrix of the original or the adjacency graph structure can be build. If not, abstractions to the graph structure have to be found that preserve or approximate information of the branching behavior.

All analyses given in this or precursory papers on search tree prediction do not include the application of transposition tables, in which visited states together with their best encountered state merits (path length plus heuristic estimate) are kept. This in fact is also a challenge for analysts. One option is the prediction of the search tree growth of IDA* with respect to bit-state hashing, which turns out to be an improvement to transposition tables in single-agent games [7] and protocol verification [3]. For this model of partial search first results on coverage prediction have been found [4].

Exact calculation of the brute-force search tree raises the question if the other source of uncertainty, namely the heuristic equilibrium distribution, can also be eliminated. As said, the equilibrium distribution of the estimate can be obtained by random sampling. However, in some cases of regular search trees exact values can be produced. If the estimate is given with respect to a pattern database storing pairs of the form (estimated value, state pattern) by analyzing the pattern database, a histogram of heuristic values can computed: we determine the number of states that satisfy a pattern with a total to be computed for each integral heuristic value in a predefined range. For consistent heuristics this range will be bounded by the heuristic estimate of the start state and the optimal solution length.

At the very far end of this research line there are precise or approximate predictions for the growth of A*'s and IDA*'s search efforts according to various kinds of heuristics, node caching strategies and problem domains. This implies an alternative way of defining heuristics themselves: ranking successor nodes according to the expected growth of the resulting search tree.

Acknowledgments I thank Richard Korf and Michael Reid for intense and helpful discussions on the topic of this paper. The work is partially supported by DFG in the project *Heuristic Search and Its Application to Protocol Validation.* Thanks to T. Lauer for proof reading.

References

1. J. C. Culberson. Sokoban is PSPACE-complete. In *Fun for Algorithms (FUN)*, pages 65–76. Carleton Scientific, 1998.
2. S. Edelkamp and R. E. Korf. The branching factor of regular search spaces. In *National Conference on Artificial Intelligence (AAAI)*, 1998. 299–304.
3. S. Edelkamp, A. L. Lafuente, and S. Leue. Protocol verification with heuristic search. In *AAAI-Spring Symposium on Model-based Validation of Intelligence*, pages 75–83, 2001.
4. S. Edelkamp and U. Meyer. Theory and practice of time-space trade-offs in memory limited search. This volume.
5. P. E. Hart, N. J. Nilsson, and B. Raphael. A formal basis for heuristic determination of minimum path cost. *IEEE Transaction on SSC*, 4:100–107, 1968.
6. M. Holzer and S. Schwoon. Assembling molecules in Atomix is hard. Technical Report 0101, Institut für Informatik, Technische Universität München, 2001.
7. F. Hüffner, S. Edelkamp, H. Fernau, and R. Niedermeier. Finding optimal solutions to Atomix. This volume.
8. A. Junghanns. *Pushing the Limits: New Developments in Single-Agent Search.* PhD thesis, University of Alberta, 1999.
9. R. E. Korf. Depth-first iterative-deepening: An optimal admissible tree search. *Artificial Intelligence*, 27(1):97–109, 1985.
10. R. E. Korf. Linear-space best-first search. *Artificial Intelligence*, 62(1):41–78, 1993.
11. R. E. Korf. Finding optimal solutions to Rubik's Cube using pattern databases. In *National Conference on Artificial Intelligence (AAAI)*, pages 700–705, 1997.
12. R. E. Korf and A. Felner. Disjoint pattern database heuristics. *Artificial Intelligence*, 2001. To appear.
13. R. E. Korf and M. Reid. Complexity analysis of admissible heuristic search. In *National Conference on Artificial Intelligence (AAAI)*, 1998. 305–310.
14. R. E. Korf, M. Reid, and S. Edelkamp. Time complexity of Iterative-Deepening-A*. *Artificial Intelligence*, 2001. To appear.
15. R. E. Korf and L. A. Taylor. Finding optimal solutions to the twenty-four puzzle. In *National Conference on Artificial Intelligence (AAAI)*, pages 1202–1207, 1996.