CONSISTENT-LABELING PROBLEMS and their ALGORITHMS

Bernard Nudel

Dept. Computer Science, Rutgers University. New Brunswick, New Jersey 08903.

ABSTRACT

Two new classes of theories have been developed giving the expected complexities of three Consistent-Labeling Problem (CLP), or Constraint-Satisfaction, algorithms: Backtracking, Forward Checking and Word-wise Forward Checking. Apart from giving the *exact* expected complexity for these algorithms for the underlying CLP distribution and domain, these theories provide useful *approximations* for the complexity of solving essentially any *individual* CLP. Given this, and the fact that the theories can reflect changes in complexity due to changes in the ordering of variables used in the search, these theories have the potential to afford significant savings for any individual CLP, by predicting, prior to search, good orderings for use in solving that CLP. We are concurrently developing improved CLP algorithms based on this and similar ordering effects.

I INTRODUCTION

The **consistent-labeling problem** (CLP) is an NP-complete problem (see [1]) of broad relevance in Artificial Intelligence and Operations Research. In [2] Haralick shows the breadth of this class of problems. (We also refer the reader there, and to [3], for references to the literature on CLP algorithms.) *Binary* CLPs, on which we concentrate, are seen to be a significant subclass, including graph and subgraph isomorphism detection, graph coloring, boolean satisfiability, and packing problems, as well as specialized cases of several other important problems. Analytically, binary CLPs provide a useful special (yet far from trivial) case from which to pursue results for general CLPs. Unless stated otherwise, the term CLP will refer to a binary CLP.

In [3] Haralick presents, for two different algorithms, an analysis of the expected complexity of obtaining *all* solutions of a CLP, for CLPs arising according to a certain one parameter probability distribution. This distribution is used by Haralick in generating random CLPs for an empirical comparison of seven different algorithms. We call these **level-0 theories**. For the CLP distribution they assume, they provide *exact* expected values.[1] However, we have found that these theories can be interpreted as providing *approximate*, but quite accurate, estimates for certain *individual* CLPs. In fact, by generalizing his analyses to apply to a more richly parameterized distribution, giving **level-1 theories**, we can obtain good estimates for *any* individual CLP solved by the corresponding algorithm (eg. fig IV-1).

A significant aspect of these generalized theories (as opposed to the level-0 theories) is that they are able to capture for a given CLP, the effects on algorithmic complexity resulting from a change in instantiation-order or consistency-checking order used during solution (eg. fig. IV-2). The theories can thus lead to significant savings by providing a basis for an intelligent choice of those orderings. We are in fact, complimentary to our mathematical analyses, developing improved CLP algorithms capable of exploiting such order-dependent effects [4, 7].

Empirically, word-wise (or bit-parallel) Forward Checking (wFC) was found best amongst the seven algorithms Haralick tested, with Forward Checking (FC) second best. Haralick obtained level-0 theories for Backtracking (BT) and Forward Checking. In [5] we obtain level-1 theories for BT and FC as well as for wFC (whose level-0 form does not appear in [3]). Level-2 theories, that improve on their level-1 counterparts, are obtained in [6] for FC and wFC; though superseded in accuracy, level-1 theories remain useful in being more manageable analytically for deriving subsequent results, in particular analytically justified ordering heuristics.

As representative of our work on FC, wFC and BT, we present here analytic and empirical results for FC only. Theory-0, theory-1 and theory-2 will be used below to denote the corresponding level theories for FC. We present theory-1, indicating its similarity to theory-2, and present experiments showing the accuracy of both theories in estimating, for individual CLPs, the complexity of solution, as well as in estimating the optimal orderings to use. A fuller account of our work is to be published in [7].

II CONSISTENT-LABELING PROBLEMS

A general consistent-labeling problem is characterized by a finite list[2] of n variables; each variable, v_i or simply i, having an associated finite domain, from which it can take any of m_i values or labels. Constraints exist on which values are mutually compatible for various subsets of the n variables. The goal is to find one or more sets of assignments of all n variables to values in their corresponding domains, such that for each assignment set all constraints are simultaneously satisfied. If the constraints exist between some *pairs* of variables, but not

[1] Apart from a subtle error that we have found, which is beside the point at this stage. See section IV.B below.

[2] We consider both the variables and their domains to be ordered.

between any triples or larger subsets of the n variables, then we have a *binary* consistent-labeling problem. Binary CLPs can be represented by their **relations matrix** $[R^{ij}_{kl}]$, a bit-matrix such that $R^{ij}_{kl} = 1$ iff the k-th value for variable i is consistent with the l-th value for variable j. Otherwise bit $R^{ij}_{kl} = 0$. Symmetries in $[R^{ij}_{kl}]$ allow us to restrict our attention to the $N \equiv \binom{n}{2}$ component relation[3] matrices $[R^{i'j'}_{kl}]$ of the relations matrix, for which $i' < j'$. For later use, we define $M_{ij} \equiv m_i m_j$, and **m** to be the vector $[m_1 \; m_2 \, .. \, m_n]$ of domain sizes.

We consider two CLPs to be **equivalent** iff their two relations matrices are identical. Then, in the class K(n **m**), of CLPs on n variables, the i-th having domain size m_i , there are $\prod_{i<j} 2^{M_{ij}}$ different CLPs. The number of ones in $[R^{i'j'}_{kl}]$ we call the **compatibility-count**, $l_{i'j'}$, for variables i' and j', and equals the number of ways that these two variables can be compatibly assigned values in their respective domains. A CLP's matrix $[l_{ij}]$, of compatibility-counts (with l_{kk} undefined), we call its **cc-matrix**; the matrix $[r_{ij}]$ of **compatibility-ratios** or simply **compatibilities**, $r_{ij} \equiv l_{ij} / M_{ij}$, we call its **cr-matrix**. Note that $r_{ij} \in \{ 0 \;\; 1/M_{ij} \;\; 2/M_{ij} \; \ldots \; 1 \}$, in contrast to their counterparts, p_{ij} , to be defined later.

CLPs in K(n **m**) having the same cc-matrix $[l_{ij}]$ (and hence cr-matrix) are **compatibility equivalent**, and fall into the same **compatibility equivalence class** or **c-class**, denoted K(n **m** $[l_{ij}]$) or K($[l_{ij}]$) when n and **m** are understood. The importance of these c-classes is explained in section III. There are $\prod_{i<j}[M_{ij} + 1]$ different c-classes for a given n and **m**, and the number of different CLPs in such a class for a given cc-matrix, $[l_{ij}]$, is $\prod_{i<j} \binom{M_{ij}}{l_{ij}}$.

Haralick [3] considers CLPs of K(n **m**), with $m_i = m$ for all i, to arise such that the probability of compatibility of any two labels for any two variables is a fixed value, p. We make use of a natural generalization of this, where a separate probability, p_{ij} , is allowed for each of the N ways to pair variables. We also generalize to allow non-equal domain sizes m_i . The corresponding probability distribution for individual CLPs, of cc-matrix $[l_{ij}]$, in K(n **m**) is then

$$p(\, \mathrm{CLP} \mid [p_{ij}] \,) \;=\; \prod_{i<j} p_{ij}^{\;l_{ij}} \,(\, 1 - p_{ij} \,)^{M_{ij} - l_{ij}}$$

Expected values of CLP-dependent quantities over all K(n **m**), when CLPs arise according to this CLP-distribution, we call **distribution averages**, as opposed to **c-class averages** which we will be using them as estimates for. How and why, are explained in the following section.

III OUR ESTIMATION STRATEGY

What we really want to achieve is GOAL A: Obtain analytic expressions for $f_k(\mathrm{CLP})$ k = 1, 2, 3, being the exact values respectively, for a *given* CLP, of its number of solutions, of nodes in its search tree and of consistency checks carried out in the search; the latter two are algorithm dependent, the last being a true indication of the algorithm complexity [3]. Exact values for any given CLP, seem an unrealistic goal. However, we have found that the spread for each of these quantities over CLPs within a c-class, is quite small; for k = 2 and 3, c-class standard deviations are about 5% to 10% of the class average values. We have thus aimed for the averages over a c-class, taking CLPs as equally likely within a class, as estimates for a given CLP of that class. If CLPs of a class are equally likely *in practice*, then this class average minimizes the sum, over CLPs of the class, of square errors from the actual values. In this sense it is the best estimate for a given CLP as a function of its $[l_{ij}]$ matrix only. In any case, because of the homogeneity amongst CLPs in a class, any distribution used would lead to essentially equally good estimates for individual CLPs in the class – the expected value of say, 5.1, 5.0 and 4.9 is near to 5.0, and hence to any of the three numbers, independent of the weights we assign to these numbers. We have then GOAL B: In order to approximate $f_k(\mathrm{CLP})$ of Goal A, obtain analytic expressions for that CLP's c-class averages

$$F_k(\, [l_{ij}] \,) \;\equiv\; \sum_{\mathrm{CLP} \in K([l_{ij}])} f_k(\mathrm{CLP}) \;\; / \;\; |\, K(\, [l_{ij}] \,) \,|$$

In [6] and [7] (for FC and wFC) we present exact expressions, constituting our level-2 theories, for these class averages. However, our chronologically earlier level-1 theories (for FC, wFC and BT) provide quite accurate approximations for them. It is in obtaining these level-1 theories that the generalized CLP generation model of the previous section provides a tool.[4] Replacing Haralick's single parameter, p, by a matrix of parameters, $[p_{ij}]$, provides sufficient structure so that the distribution average

$$E(\, f_k(\mathrm{CLP}) \mid [p_{ij}] \,) \;\equiv\; \sum_{\mathrm{CLP} \in K(\, n \;\, m \,)} f_k(\mathrm{CLP}) \; p(\, \mathrm{CLP} \mid [p_{ij}] \,)$$

of a CLP-dependent quantity, $f_k(\mathrm{CLP})$, can provide a good estimate of that quantity's c-class averaged value, $F_k(\, [l_{ij}] \,)$, for *any* given c-class; providing we know how to choose the distribution parameters, $[p_{ij}]$, appropriately for that class. In [5] we present several reasons why a good parameter-matrix choice for approximating a class average $F(\, [l_{ij}] \,)$, is $[p_{ij}] = [r_{ij}] \equiv [\, l_{ij} / M_{ij} \,]$. Given this, we then have our final GOAL C: Obtain analytic expressions for the distribution averages $E(\, f_k(\mathrm{CLP}) \mid [p_{ij}] \,)$. Then as approximations for the c-class averages $F_k(\, [l_{ij}] \,)$, of Goal B, use $E(\, f_k(\mathrm{CLP}) \mid [r_{ij}] \,)$. Haralick's level-0 results provide these distribution averages for the $[p_{ij}] = [p]$ case.[5] We have found them good estimates for averages over c-classes having uniform $[r_{ij}] = [r]$, when p = r is used in his level-0 theories. We generalized to include non-uniform $[p_{ij}]$ so as to obtain level-1 theories that are equally accurate for classes with a non-uniform cr-matrix $[r_{ij}]$.

[3] Primed indices denote fixed though unspecified values.

[4] Here, unlike in [3] where it models an actual experiment, it is a technical device only, for giving distribution averages that approximate c-class averages. Its use for this purpose implies no assumptions about the CLP distribution occuring in practice.

[5] We use [x] to indicate a matrix all of whose (defined) elements are equal to x.

IV FORWARD CHECKING - EXPECTED COMPLEXITY

A. Nomenclature

The **instantiation order**, $F_0 \equiv (i_1 \;\; i_2 \;\; \dots \;\; i_n)$, denotes the *globally fixed*[6] order in which the n problem variables are **assigned** a value during the search. All nodes at level k of the tree have the same **assigned variables** $A_k \equiv (i_1 \; i_2 \dots \; i_{k-1} \; i_k)$. Variables that remain to be instantiated after level k, make up the list of **future variables** for that level $F_k \equiv (i_{k+1} \; i_{k+2} \dots \; i_n)$. For added generality we allow an independently specifiable **consistency-checking order** $G_0 \equiv (i'_1 \, i'_2 \dots \; i'_n)$, according to which, future variables $f \in F_k$ are to be selected for forward checking at any node of the search tree; at level k they are chosen in the order of the list G_k , defined as F_k reordered according to G_0. The j-th element of the list G_k is denoted g_{jk} . In the theory below, products and sums over zero number of terms are to be taken as resulting in 1 and 0 respectively.

B. Analytic Results

With the above definitions, we now present our level-1 theory for FC (theory-1), giving the *distribution averages* of f_k(CLP) k = 1, 2, 3. (See [5] for the derivation by a direct generalization of that in [3]. In [6] and [7] an alternative recursive derivation is given that has the advantage of unifying both level-1 and level-2 theories.) The expected number of solutions is S of eqn. (1). This is not a function of the algorithm used but other results here *are* specific to the FC algorithm. The **expected number of nodes that appear at level k of the FC algorithm's search tree** is n_k of eqn. (2). $S_f^{(k)} = 1 - W_f^{(k)}$ is the **Survival probability** of variable f when the first k variables of F_0 have been instantiated, expressed in terms of the f variable's **Wipe-out probability**: $W_f^{(k)} = (1 - \prod_{i \in A_k} p_{fi})^{m_f}$.

$$S \;=\; (\prod_i m_i) (\prod_{i<j} p_{ij}) \tag{1}$$

$$n_k \;=\; (\prod_{i \in A_k} m_i) (\prod_{i<j \in A_k} p_{ij}) (\prod_{f \in F_k} S_f^{(k-1)}) \tag{2}$$

$$m_k^f \;=\; m_f (\prod_{j \in A_{k-1}} p_{fj}) \,/\, S_f^{(k-1)} \tag{3}$$

$$c_k \;=\; \sum_{i=1}^{|G_k|} (m_k^{g_{ik}} \prod_{j=1}^{i-1} S_{g_{jk}}^{(k)} \,/\, S_{g_{jk}}^{(k-1)}) \tag{4}$$

$$C \;=\; \sum_{k=1}^{n} C_k \;=\; \sum_{k=1}^{n} c_k n_k \tag{5}$$

The **expected number of labels for future variable f**, *before* **forward checking at level k** is m_k^f of eqn. (3), in terms of which the **expected number of** consistency-checks **per node, during forward checking at the k-th level**, is c_k of eqn. (4). The **expected (total) number of consistency checks during forward checking at the k-th level** is then given by: $C_k = n_k \, c_k$. The sum, C, of C_k over the n levels of the tree is an appropriate unit for the expected complexity of FC [3].

The relation between Haralick's level-0 theory for FC and our level-1 theory above requires that our results reduce to his when all p_{ij} = p and all m_i = m. This is in fact the case with the exception of expression (4) for c_k. This difference is due to a subtle error we have found in [3], so that Haralick's c_k is the specialized form of eqn. (4) with the denominator term, $S_{g_{jk}}^{(k-1)}$, removed (i.e. replaced by 1). The theory-0 used below is the corrected version, using the specialized form of eqn. (4) as given here.

Though our level-2 theory for FC [6, 7] requires a more complicated expression for the wipe-out probability $W_f^{(k)}$, we note here that apart from this, the more accurate level-2 theory is exactly as for the level-1 theory above (but with p_{ij} already replaced throughout by r_{ij}, as we are required to do in fact, when applying theory-1 for a class with cr-matrix $[r_{ij}]$).

C. Examples and Discussion

Remember that the above theory-1 gives distribution-average values for CLPs arising according to the generalized model of section II (and as such they are exact), and these are to be used as estimates for c-class averages by setting $[p_{ij}]$ equal to the class cr-matrix $[r_{ij}]$. Theory-2 gives these c-class average values *exactly*. A c-class average in turn, is intended to be a useful estimate for any *individual* CLP of that class. The following experiments show the usefulness of both our theory-1 and theory-2 for these ends.

Figure IV-1 compares theory-1 and theory-2 with experiments over all 125 c-classes for the 4096 CLPs with n = 3 and m_i = 2 ∀ i, using orderings $F_0 = G_0 = (1 \; 2 \; 3)$. For each c-class, all CLPs were solved by FC. The experimental c-class average of the number of nodes/CLP (#nodes) and consistency-checks/CLP (#checks) in a problem's search tree is shown. For the latter (only) we also indicate the corresponding experimental standard deviations for the c-classes (note that they are quite small and many are even zero!). As required, we use in theory-1 $[p_{ij}] = [l_{ij} \,/\, M_{ij}]$ for each of the 125 classes arising from $l_{12}, l_{13}, l_{23} \in \{ 0 \; 1 \; 2 \; 3 \; 4 \}$.

Figure IV-2 shows, using the 5-Queens problem and another CLP randomly selected from the same c-class, the ability of theory-1 and theory-2 to reflect complexity changes resulting from changes in instantiation and consistency-check order (here $F_0 = G_0$) used in the algorithm. The permutation# indicates the permutation of (1 2 3 4 5) used for these orders, where permutations are arranged in lexographic (increasing numerical) order; the 60-th being (3 2 5 4 1).[7] As required, we set $[p_{ij}]$ to the problems' (common) $[r_{ij}]$ matrix in applying theory-1. It is possible to show generally that for n-Queens problems,

$$[r_{ij}] \;=\; [\; (n^2 - 3n + 2 | \; i-j \; | \;) \,/\, n^2 \;]$$

[6] The algorithms we are developing have orderings that are dynamically decidable at each node. However a globally fixed order·acts as default in case of decision ties. It is a good choice for this default that we hope to arrive at through our analyses.

[7] Results for theory-0, theory-1, theory-2 and the 5-queens CLP are symmetric, and for CLP_2 nearly symmetric, about "permutation# 60.5". We therefore display, and discuss, results only up to permutation# 60.

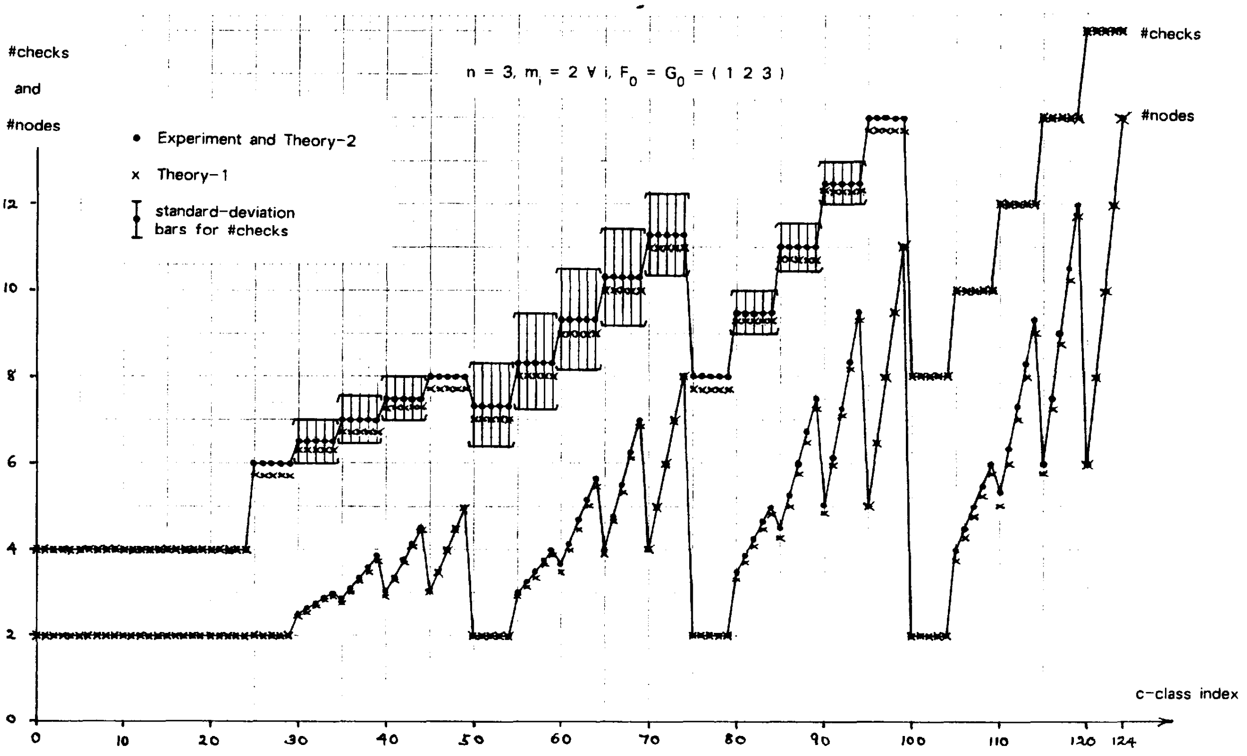

Figure IV-1: Experimental and theoretical c-class averages for CLPs of K(3 [2 2 2]), using FC.

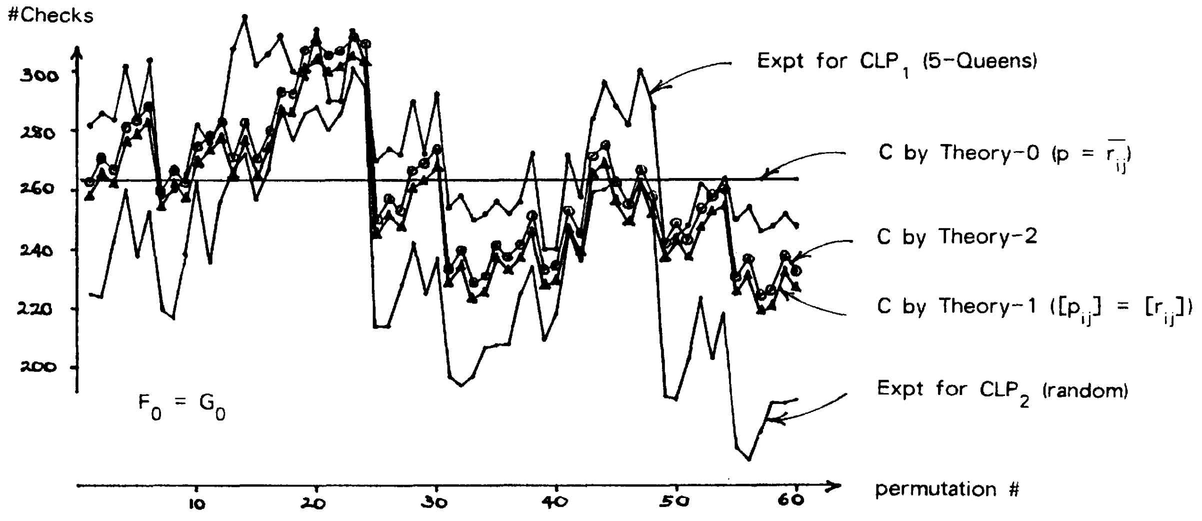

Figure IV-2: Variation of the #checks in the FC search tree, with variables-ordering used, for CLPs in the 5-queens' c-class.

The figure shows empirical results for two *individual* CLPs representative of their c-class, in comparison with the theoretical results which give *class averages*. Agreement between theories 1 and 2 and the two individual CLPs is seen to be good, especially in regard to ranking of orders. For example, the permutation that is optimal for reducing average complexity over the c-class, is given by the exact theory-2 to be permutation# 57 or (3 2 4 1 5). The approximate theory-1 is seen to agree. Moreover, for the individual CLPs this is in fact the optimum order for CLP_2, and the 5-th best for use with CLP_1 (5-queens).

Regarding predicting class averages: For the class average number of solutions, theory-1 gives the same, exact result as theory-2. For #nodes and #checks in the search tree, the theory-1 approximation seems to consistently underestimate slightly the exact theory-2 c-class averages. This underestimation varies with c-class and order(s) used, but seems to generally be small: about 5% to 10% (*less* than 5% everywhere in figs. IV-1 and IV-2). Furthermore, this underestimation needn't impare theory-1's ability to predict optimal (w.r.t class average) orderings; as long as theory-1 "shadows" well the variation of the exact theory-2, so that the peaks and troughs of the two occur essentially at the same orderings; Figure IV-2 and similar plots certainly show this to be the case.

Regarding predicting values for individual CLPs using the class average: In preliminary experiments we have found that for the #nodes and #checks in a tree, the ratio of standard deviation over a c-class to the class average value is also around 5% to 10%. This is reflected in the standard deviations in fig. IV-1 and by the closeness of the two CLP curves, to the exact theory-2 curve in fig. IV-2. Our partitioning scheme is thus justified as being appropriate for our aim of obtaining quite homogeneous classes so that a c-class average, in general, gives a good estimate for *individual* CLPs of the class.

V IMPROVED CLP ALGORITHMS

The ability, shown for FC in fig. IV-2, of our level-1 and level-2 theories to capture the order-dependence of complexity for the three algorithms analysed, suggests using the theories in determining (quasi) optimal global orderings before search begins. This requires solving, or inverting, the theories to express the complexity-minimizing orders as a function of cc-matrix, rather than giving complexity as a function of orders and cc-matrix, as at present. We have not yet achieved such an inversion. However, there are several promising paths to at least an approximation method for such optimum global orderings. Until recently, we have concentrated on a simpler goal of using the theories and heuristic reasoning to suggest *locally* good orderings. These are built up of a sequence of choices that are individually good without ensuring the goodness of the overall orders. The theory can be used for this because each node of a tree corresponds to sub-CLP of the original CLP; a k-th level node corresponds to a sub-CLP having n-k variables, those in F_k at the node, with respective domains being the values that have survived from the original domains for those variables. As a result, any local heuristics suggested by the theory for use at the root can also be applied at any node throughout the tree, using the new parameters for that node. We have obtained and tested several such heuristics for instantiation ordering and consistency-check ordering. Results are presented in [4, 7]. Such *dynamically-determined* orderings have resulted in significant improvements. However, further efficiencies exist to be had using a combined local/global (dynamic/static) approach where global orderings provide good defaults when local heuristics lead to ties. Theoretical determination of good global-ordering heuristics is our next goal.

ACKNOWLEDGMENTS: This work has benefited greatly from the rigorous standards of my supervisor, William Steiger, and from the encouragement of Saul Amarel (in whose course this work was first started) and Marvin Paull.

1. Haralick, R. M., Davis, L. S. and Rosenfeld, A. "Reduction operations for constraint satisfaction." *Information Sciences 14* (1978), 199-219.

2. Haralick, R. M. and Shapiro, L. G. "The consistent labeling problem: Part I." *IEEE Trans. Pattern Analysis and Machine Intelligence PAMI-1*, 2 (1979), 173-184.

3. Haralick, R. M. and Elliot, G.L. "Increasing tree search efficiency for constraint satisfaction problems." *Artificial Intelligence 14* (1980), 263-313.

4. Nudel, B. Improved constraint satisfaction algorithms using inter-variable compatibilities. Tech. Rept. DCS-TR-108, Computer Science Dept., Rutgers University, Dec., 1981.

5. Nudel, B. Consistent-labeling problems and their algorithms - Part I. Tech. Rept. DCS-TR-112, Computer Science Dept., Rutgers University, 1982. To appear.

6. Nudel, B. Consistent-labeling problems and their algorithms - Part II. Tech. Rept. DSC-TR-113, Computer Science Dept., Rutgers University, 1982. To appear.

7. Nudel, B. "Consistent-labeling problems and their algorithms." *Artificial Intelligence 20* (1983). Special issue on Search and Heuristics, in memory of John Gaschnig; to appear.