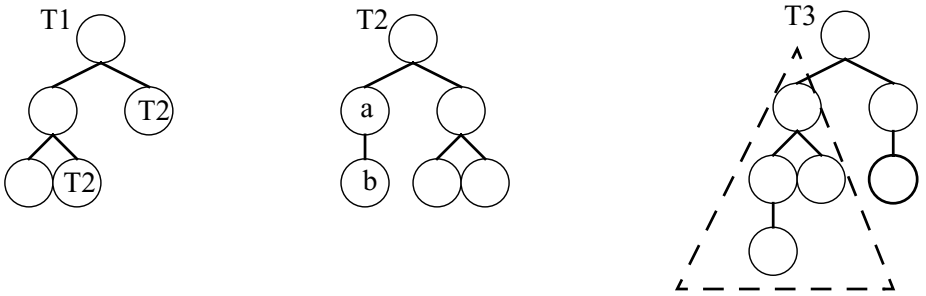
ACGP is a new method to explore regularity

Cezary Z. Janikow
Department of Math and Computer Science
University of Missouri - St. Louis
cjanikow@ola.cs.umsl.edu

GP inherently is a technique for, and relies upon, exploration and exploitation of regularity. With ADFs, regularity translates into modularity. Regularity is explored both implicitly (evolution via crossover) and explicitly (as in ADFs).

ACGP (Adaptable Constrained GP) offers a new way for GP to explore regularity of a different kind. It will have an impact on modularity as well. I will here briefly describe *what* kind of regularity ACGP explores, and how it differs from ADFs. As to *how* this is done, I will be presenting this during my regular presentation (ACGP was developed for tree-based chromosomes, so this discussion is on trees).

GP solves problems by implicitly combining Building Blocks via evolution (same as GA, just different kind of BBs). BBs don't have to be complete subtrees, they are simply tree fragments, with possible subtrees represented by a single "don't care" node. When it comes to explicit exploration of regularity, such as leading to modularity by ADFs, the regular structures are complete subtrees, that is every leaf must be a terminal instance (or an ADF, which in fact relaxes the terminal requirement - I have to admit I need to catch up on your recent work). Whether ADFs are complete subtrees or not, they change the alphabet (by adding ADFs to the set of terminals), but not the probability distribution of crossover and mutation. In other words, they explore regularity among BBs, or rather subtree instances, but not regularity in the probability distribution. The latter is what ACGP does.



Another way of looking at this is to say that ADFs in fact change the probability distribution, in a more indirect way. For example, suppose the above T2 is an ADF, and it is used as shown in T1. Then, the specific regular BB a-b will appear in T1 more frequently than by a chance, thus in fact the effective distribution of different BBs becomes non-uniform.

ACGP explores this distribution space more directly. Now assume that T3 is subject to crossover, to replace the indicated subtree, and that the source parent is T2 (a chromosome, not ADF). There are 6 potential subtrees that can be crossed to T3: three of them are just leaves, two are subtrees, and one complete tree. Under normal scenarios, disregarding any internal/external preferences, each of the six will have the same probability. Now assume that somehow the choices are restricted, as through STGP, CGP, CFG-GP, or some other technology. And suppose that only three are now plausible. These three still have uniform probability of being used. Thus the space of possible crossovers has been pruned, but the remaining choices are still non-distinguishable. But now suppose that T3 will be better off if, out of the three choices, the subtree a-b becomes the root's left child. Can we learn this information, and use it to adjust the probability distribution more directly? This is what ACGP does. So in fact, it discovers regularities in the probabilistic representation space rather than in the space of tree instances. At present, it only works at the first-order level of parent-child but we plan to eventually extend this.