**3**

**Genetic Algorithm**

**3.1. Primeiro nível de subtítulos em Arial 11pt negrito minúsculas pequenas**

A genetic algorithm (GA) is an optimisation method inspired in nature and its evolution mechanisms. Starting with a set of typically randomly generated solutions – called individuals of the initial population - the algorithm tries to breed new individuals based on the previous generation using operators of selection, mutation and crossover similar to those used by biological organisms.

This is a metaheuristic method and, as such, the solution space to be explored is limited by various constraints such as computational capacity, algorithm parameters, problem information and time constraints. Give this limitation the optimum returned by the algorithm may not be the global optimum but a satisfactory solution for the problem, a local optimum.

A GA generally works in the following way:

The initial population is generated and populated with individuals whose chromosome is randomly generated, each individual represents a different solution for the problem.

A fitness function runs through every individual and awards a fitness value to each solution, it represents the merit each solution has and is used further down the algorithm to select parents for the next generation. A fitness function can be static o dynamic, changing at runtime.

When every individual of the initial population has a fitness, value assigned, an Evolve command is issued triggering a set of operations.

The parents of each new individual are selected by a method that takes into account the fitness value of each individual of the current population, individuals with better fitness are more likely to be selected. The selection is passed on to the first genetic operator, usually a crossover function that has the task of assembling a new chromosome made up of blocks from the parents’ chromosomes. By this point the chromosome of the new individual is built, however, to preserve the diversity of solutions during the search of the optimum a mutation operator is used the introduce new information in the chromosome.

The new individual is then evaluated with the fitness function, and the process is repeated until a predefined stop criteria is met.

[fazer imagem bonita em UML do ciclo todo]

The next subchapters will detail the fundamental elements of a GA.

**3.2. Fundamental elements of a genetic algorithm**

3.2.1. Individual

In nature, a set of organisms from the same species defines a population and each organism is defined by a set of encoded instructions on its DNA. Each chromosome sets specific traits such as eye color, and each trait can have multiple values: blue, brown, green, etc.

An individual in a GA does not represent an organism but a possible solution for the optimisation problem. The way it solves the problem is encoded in a convenient datatype where each characteristic of the solution is stored, similar to DNA in natural organisms. In general, an individual shares with the others the structure of its chromosome, but the values stored change from individual to individual.

[imagem olhos cabelo etc etc]

The original representation of a chromosome is a simple bit string vector – a series of 0 and 1 – but even though it is the classical representation it is not always the most useful. For different problems, other datatypes can be better suited to store the characteristics of each individual, for example, a given problem might be better solved using individuals where the DNA is stored in a multi-dimensional array and where the values stored do not use the base 2 (binary) numeric system.

The datatype choice is problem dependant, the number of different variables that are changed during the optimisation and their relationship with each other can exclude certain datatypes and make others more attractive for the developer. As an example, when optimising a structure’s sections a vector can be used to represent the section in each bar, but if in addition to the section, the node coordinates (x,y,z) also need to be optimised an array is probably the best way forward to represent the DNA of each individual solution. [imagem]

[imagem de caixas binário] [imagem dos data types falados]

3.2.2. Initial population

The initial population is the beginning of the GA, if there is previous knowledge of where the optimal solution is located in the search space the initial population can be seeded near that solution space to save time in the search. When there is no information that points the developer in a specific direction to where the ideal solution is located a random set of individuals is generated to make the initial population.

[IMAGEM SOLUTION SPACE PROPAGADO]?

The size of the initial population usually increases with the number of genes the GA operates with in the search of the optimum. Using the example in 3.2.1, the GA that only optimises the bar sections will need a smaller population size that the second GA that optimises the section and topology of the structure, in order to populate the solution space as well as the first GA.

3.2.3. Fitness function

The Fitness function is a critical part of the GA, it is responsible for the merit evaluation of each solution so that the best individuals have a larger probability to pass on their genes. To do that successfully it needs to evaluate each solution and also give penalties for bad solutions to decrease reproduction probability.

In simple problems, the definition of the fitness function can be straightforward, when there is a single and clear objective for the optimisation, for example:

-Arrange various shapes to be cut in a fabric of WxH in order to reduce material waste. Fitness function: Evaluate the wasted area of each solution, apply penalties when the shapes overlap.

For more complex problems the first challenge arises, optimise more than one parameter.

When more than one parameter is evaluated by the fitness function there needs to be a way to translate that multiparameter evaluation in a single classification that represents the quality of the solution, that operation can be quite difficult, for example:

When optimising a structure in addition to the material usage the might be specific nodes in the structure that cannot deflect more than a given value, several questions can be made:

* How to weight the rigidity and mass parameters of the structure?
* When check deflection is every measure node equally important?
* How to add a penalty for a structure with good rigidity and poor material usage?
* Is the penalty so harsh that it can turn the algorithm into a random search?

The second challenge, is very common in engineering problems, the impossibility to evaluate directly each solution, most Gas applied to engineering problems will need to evaluate each solution in FEM of CFD software and retrieve the results to evaluate the solution, for example: To optimise the drag coefficient of a car a call to an external CFD program needs to be made to get the drag coefficient of each solution. This need for external calls to complex software packages can significantly delay the GA making computational power and time an important constraint in such problems.

The fitness evaluation can be static or it can vary at runtime as the generations increase, that change in evaluation is useful in very sensitive problems where the scale of penalties awarded initially to arrive at a near optimum are inadequate for the final adjustments needed to arrive at the optimum.

When it is impossible to define, a fitness function the adoption of interactive GAs might be an appropriate solution depending on problem complexity. In interactive GAs the fitness function is the user that is asked to rate each individual, it is a useful approach when the algorithm is trying to optimise subjective based on a subjective parameter such as aesthetics.

3.2.4. Selection function

The selection function is used in conjunction with the fitness function to translate the Darwinian concept of survival of the fittest to the optimisation algorithm.

There are various ways to implement the selection function, the most commonly used are the roulette wheel selection and tournament selection, both are explained in the next subchapters. Even though there are clear differences between the various possible algorithms, the final goal is the same, ensure the individuals with better fitness classification are selected more often so that their genes are carried over to the next generation. When repeated over generations, it acts as a filter for the bad genes.

3.2.4.1. Roulette wheel

The roulette wheel algorithm is the most commonly used method in traditional GAs. The key operations are:

* Sort the individuals according to their fitness value;
* Add the fitness values of each individual in the population and store the value:
* Assign to each individual a selection probability:
* Randomly generate a number between 0.0 and 1.0;
* Sum the probability of each individual until the generated number is smaller than that sum;
* The last individual that was added is the one selected to breed.

The implementation is straightforward and is effective in problems where convergence is easily attainable. When there is a need for more control on the selection function the tournament selection is a better option.

[imagem da roulette wheel]

3.2.4.2. Tournament selection

This method as the name implies is based on a tournament between individuals. A pool of random individuals is created, and their finesses are compared, that tournament can return a best individual or a probability can be set for the selection of the best individuals, for example: the best individual is selected 70% for the time, the second 20% and the third 10%.

By having two adjustable parameters it is possible to fine tune the selection algorithm to better suit the problem. An increase in the pool of individuals increases the selection pressure, decreasing the probability of the worse individuals being selected and by changing the probability of selection among the best individuals, early convergence is avoided.

[meter uma imagem.. fica bonito]

3.2.5. Genetic operators

Genetic operators receive the chromosomes from the parent individuals, and with manipulation of genes return a new individual inspired on its parents. In many cases the crossover operator is already and acceptable way of solving a problem using GAs however, it is common to add a mutation operator to ensure the diversity of the solutions. (Fogel 2006 “Evolutionary Computation: Toward a New Philosophy of Machine Intelligence”)

3.2.5.1. Crossover

The crossover operator receives de gene blocks from the parents and assembles them to create a new chromosome made of gene blocks from each parent that is assigned to the new individual.

There are several methods to create a new chromosome, the most common are single and double point crossover that will be detailed next. The uniform crossover is another method that widens the search space of the GA, because of this property it is only suited for simple problems where the starting search space is not too large or when constrains such as computational capacity and time are not relevant.

The single and double point methods work in the same way, a the chromosome of one parent is cut at a specific point and from that point onwards the chromosome is replaced the the genes from the other parent. The difference between single and double point is in the final point of mutation, on the single point crossover the genome uses genes from the second parent until the end of the genetic code, for the double point crossover the end point is calculated and the replacement of genes is conducted until that point. [meter a figura]

3.2.5.1. Mutation

This operator adds new information to the chromosome that is received from the crossover function, t

This operator changes randomly information on some genes of the chromosome that is received from the crossover function, the objective is to preserve diversity in the solution space so that early convergence in a local optimum is not likely.

The operation is defined as a probability that each gene has of being changed to a random value, that probability needs to be low enough to ensure the search doesn’t degrade into a random search of the optimum.

The way the mutation is achieved depends on the datatype used to store the chromosome, for the classic bit string format the mutation is achieved by a simple random change from 0 to 1 in some places of the string, as seen on the image. [meter figura]

Mutation when the gene values are not in base 2 (binary) needs to be carried with extra caution to avoid excessively drastic value changes that are particularly bad in the final phases of the GA when the fine tune of the solution is needed. To have that extra control a method like the Gaussian mutation with the mean in the current value of the gene is desirable, this was the algorithm used in the present work and will be detailed in chapter 4.

3.2.4. Termination criteria

There needs to be a way of telling the algorithm to stop the search, that instruction is issued by the termination criteria. The developer has some leeway when defining this block of code but some circumstances might need to be considered, termination examples are:

* Maximum number of generations;
* Found a sufficiently good solution;
* No recorded improvement over the last X generations;
* Computational time limit exceeded.

**3.3. Problems of genetic algorithms**

Meto isto aqui ou no inicio/introdução????