

3

Algoritmo Genético e a Reconstituição de Acidentes

Métodos de otimização são utilizados quando se pretende minimizar ou maximizar uma função. Estes algoritmos procuram, de formas diferenciadas, chegar a valores ótimos. De forma geral, todos se utilizam de iterações, comparando os valores das anteriores com as novas, e através de algumas regras de decisão, escolhendo por qual “caminho” seguir adiante.

O Algoritmo Genético (GA) não trabalha diretamente com o problema a ser otimizado. As equações que regem o sistema não são trabalhadas neste otimizador. Esse algoritmo gera inúmeros conjuntos de valores aleatoriamente, e uma função de avaliação dirá quão bom um dado conjunto é em relação aos demais. Feito isso, a nova geração de valores se dará pela combinação dos antigos, sendo que os mais bem avaliados pela função terão maiores probabilidades de serem “escolhidos” para se misturarem. Para efetuar estes procedimentos existem diversas ferramentas que de formas diferenciadas têm como objetivo direcionar esta evolução para os valores ótimos.

3.1.

Algoritmo Genético

Na escolha do método a ser utilizado dois fatores foram levados em conta. O primeiro é quanto à robustez do algoritmo, ou seja, ao se alterar o programa de simulação, quantas seriam as mudanças necessárias no algoritmo de otimização. Esta característica foi levada em conta, pois o programa de simulação desenvolvido pelo grupo de pesquisas de Dinâmica de Veículos da PUC-Rio ainda está em fase de aperfeiçoamento. Entre as mudanças previstas, está a inclusão de um modelo de veículo flexível. O segundo fator foi o estado da arte, que mostra uma tendência na tentativa de utilização da computação evolucionária em diversas áreas como suporte à decisão, classificação de dados, reconhecimento de padrões, previsão, otimização, controle, modelagem, planejamento, descoberta de conhecimento, entre outros. A consideração de tais fatores levou à escolha do Algoritmo Genético.

O Algoritmo Genético é caracterizado pela geração de valores aleatórios segundo diferentes distribuições estatísticas e pela evolução desses valores, o que se dá de forma análoga à evolução dos seres vivos de acordo com a Teoria Evolutiva de Darwin. Esta prevê que os indivíduos mais aptos têm maiores chances de sobrevivência. Como na natureza, em alguns casos de engenharia são muitos os fatores que definem a aptidão de um “indivíduo”, tal como definido a seguir. Portanto o cruzamento dos mais aptos com aqueles com menos chance de sobrevivência também é importante, para que se tenha maior diversidade de informações e portanto maiores chances de se chegar ao “indivíduo mais apto”. No caso estudado, o número de informações é tamanho que a análise separada de cada parâmetro relevante à sobrevivência e de suas interações torna-se impraticável. Portanto, a sinergia proveniente do conjunto das informações criadas em um primeiro momento de forma aleatória, direcionada pelas avaliações de cada grupo de informações, torna-se um recurso eficaz de busca do ótimo sem que seja necessário recorrer a uma pesquisa exaustiva da influência de cada parâmetro.

Alguns dos principais conceitos aplicados em algoritmos genéticos são:

Gene – Valor a ser otimizado.

Indivíduo ou Cromossomo – Vetor com todos os parâmetros que se pretendem otimizar. Este vetor será manipulado através dos operadores.

População – Grupo de indivíduos. Denomina-se população inicial a primeira população. Esta tem uma particularidade importante quando comparada às seguintes: a criação de seus indivíduos é dada de forma que cada um não tem nenhuma ligação com os demais a não ser por suas distribuições de probabilidade.

Tamanho da População – Número de indivíduos de cada geração. Quanto maior o tamanho da população, maior tende ser a diversidade genética, porém maior o tempo gasto em avaliação e cruzamento dos indivíduos a cada geração.

Geração – O momento em que se encontra a evolução. A população da primeira geração dá origem às demais e a princípio é gerada aleatoriamente. Pode-se utilizar outras distribuições em sua criação, mas o habitual será a distribuição aleatória. As subseqüentes são geradas principalmente pelo cruzamento das informações das anteriores.

Função Avaliação – Esta é responsável por dar uma “nota” (aptidão) para um cromossomo. Dada esta aptidão, as gerações seguintes serão formadas, sendo que os mais aptos terão maior probabilidade de perpetuar parte do seu cromossomo do que os menos aptos.

Seleção – Critério utilizado para escolher quais os indivíduos que serão “misturados” através do operador de cruzamento (a seguir). A seleção é baseada na avaliação dada pela função descrita anteriormente e tem influência direta sobre a pressão evolutiva, pois pode ser configurada de forma a diferenciar fortemente os indivíduos mais aptos ou tornar menos improvável o cruzamento dos menos aptos.

Cruzamento (*Cross-Over*) – Pares de indivíduos são escolhidos aleatoriamente pelo critério descrito anteriormente e os valores dos seus vetores são combinados por este operador de forma a gerar novos indivíduos para a nova geração. Esta operação tende portanto a combinar os indivíduos mais aptos, convergindo assim a valores ótimos.

Mutação – Este operador troca aleatoriamente um ou mais valores de um ou mais indivíduos escolhidos aleatoriamente. Esta operação é responsável pela diversificação genética da população e é importante principalmente para casos em que o algoritmo esteja tendendo a um ponto de ótimo local e não global.

Elitismo e *Steady State* – No primeiro caso o melhor indivíduo, e no segundo caso os melhores indivíduos (quantidade definida pelo usuário) da geração anterior são mantidos inalterados para a geração seguinte. O resto da população seguinte é o produto de evolução a partir da anterior através dos operadores já discutidos.

Critério de Parada – Pode-se estipular um valor mínimo (ou máximo) na função de avaliação o qual define a decisão de parada. Demarca-se muito comumente também o número de gerações que serão utilizadas ou o tempo limite de processamento. Outros critérios de parada podem ser definidos.

Soluções Inválidas – Soluções fora das restrições do problema. Este tipo de solução pode ser trabalhado através de penalidades, porém deve ser fortemente evitado, pois gasta tempo de computação e quando em grande número, pode resultar em soluções não ótimas.

3.2.

Algoritmo Genético Aplicado à Reconstituição de Acidentes

Dadas as definições do subitem 3.1, são estabelecidos agora os critérios para uma “cooperação” adequada entre o otimizador e o simulador. Um fator importante a ser definido é a função de avaliação, pois esta será a conexão responsável por “direcionar” o algoritmo para os melhores indivíduos. Quanto melhor a avaliação de um indivíduo, maior probabilidade ele terá de passar as suas informações para as gerações seguintes. Portanto, ao se escolher a função, deve-se ater tanto à preocupação de bem diferenciar entre um bom indivíduo e um mau indivíduo, quanto à de simplicidade em sua formulação, pois uma excessiva complexidade aumentaria o risco de se formularem hipóteses inválidas. Partindo destes dois princípios optou-se pela raiz quadrada da soma dos quadrados das distâncias entre as posições obtidas e esperadas dos veículos. Este cálculo é efetuado tomando-se a posição, após a parada, de dois pontos da carroceria de cada veículo (de preferência que sejam geometricamente opostos) e medindo-se as distâncias com relação às respectivas posições esperadas, como se pode ver nas Figuras 22 e 23 (f).

Um outro desafio é definir quais serão as variáveis que o algoritmo deverá trabalhar em busca da solução ótima. Quanto mais fatores a se otimizar existirem, mais amplo será o universo de busca do problema e portanto mais lentamente se dará a convergência ao ponto ótimo. Por outro lado, se os fatores estiverem demasiadamente presos a limites que não sejam necessariamente condizentes com a realidade dos acontecimentos, mais provavelmente não se chegará a uma solução válida. Por estes motivos definem-se, a princípio, as entradas (genes) do algoritmo como:

- Velocidades iniciais;
- Posições iniciais;
- Atitudes iniciais;
- Local da colisão;
- Partes colididas dos veículos;

Os limites e/ou a média e distribuição probabilística destes fatores serão dados de entrada do usuário/especialista. Com isso, apesar do grande número de parâmetros a ser otimizado, o universo de busca será limitado. Quanto mais informações reais o analista tiver, será de se esperar uma convergência mais rápida do algoritmo ao ótimo global.

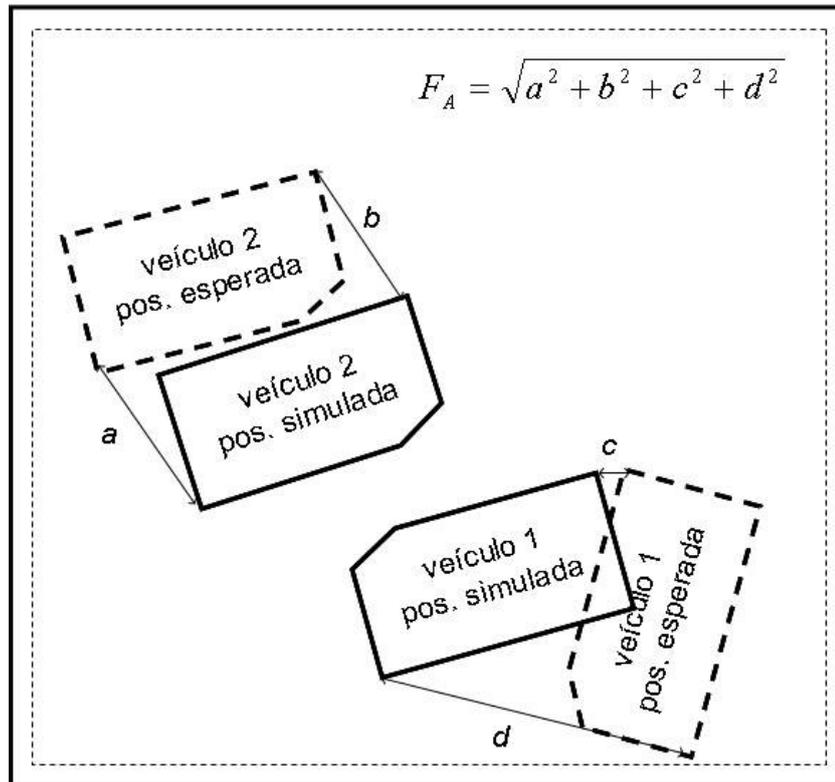


Figura 21- Função de Avaliação

Estes limites serão respeitados no momento de criação da população inicial e serão utilizados operadores que também respeitem este limite. Por estes motivos, será recomendado ao usuário que utilize limites razoáveis, no sentido de evitar a escolha de uma faixa excessivamente estreita, a ponto de deixar marginalizadas as soluções ótimas.

As entradas do algoritmo não precisam ser necessariamente nas grandezas dadas. Pode-se por exemplo, parametrizar todos os genes do cromossomo para valores entre 0 e 1, como se pode ver na Figura 23 (a), com quantas casas decimais se façam necessárias e com as devidas distribuições probabilísticas (definidas pelo usuário). Num segundo momento, já independente do otimizador, transformam-se estes valores em equivalentes entre os limites inferior e superior (também estabelecidos pelo analista), como mostrado na Figura 23 (b).

Definir a geometria do problema talvez tenha sido o maior desafio. Como o algoritmo genético é baseado em um processo estocástico, as posições dos veículos não podem ser definidas em coordenadas globais. Esta impossibilidade é devida ao fato de que inicialmente o otimizador gera valores independentes entre si. Assim, mesmo que fossem impostos limites para cada uma das variáveis de localização de cada veículo, seriam no entanto muito grandes as

chances de geração de situações onde os veículos não colidiriam, ou colidiriam de forma a não respeitar as áreas de contato pressupostas. Seriam portanto geradas inúmeras soluções inválidas, que por sua vez “gastariam” muito tempo computacional e poderiam até levar ao paradoxo de situações irreais com valor da função de avaliação típico de uma solução ótima, como por exemplo uma configuração onde os veículos já se encontrariam perto das posições finais e nem chegariam a colidir.

Uma solução encontrada para este problema foi definir a posição inicial de cada veículo através três parâmetros. O primeiro seria o ângulo formado entre o sistema de coordenadas locais do veículo, com origem no centro geométrico do mesmo, e o sistema de coordenadas globais. Os demais parâmetros seriam o local de colisão e a parte colidida de cada veículo (Figura 23 (c)). De acordo com o modelo de colisão instantânea descrito no capítulo 2, o ângulo que cada veículo forma com o sistema de coordenadas globais é obtido a partir das definições das áreas afetadas na colisão. Com isso, o algoritmo genético irá gerar valores iniciais para as partes colididas e para o local de colisão, respeitando os limites estabelecidos, sem criar portanto soluções inválidas a priori.

Através destas informações, é possível conhecer a posição global de cada um dos veículos a cada variação dos parâmetros pelo algoritmo. Basta girá-los pelo ângulo, considerar as partes colididas e ancorá-las no local de colisão (Figura 23 (d)). Cabe ressaltar que cada uma das características acima (parte colidida e local de colisão) será aleatoriamente variada pelo GA, dentre os limites estipulados, sem que soluções inválidas sejam geradas. É importante destacar que o procedimento de determinação da posição global será todo efetuado no programa de simulação, portanto à parte do algoritmo otimizador.

Observa-se que a Figura 23 representa conceitualmente o problema que está se tratando, uma vez que pequenas modificações foram feitas no intuito de simplificar a sua implementação. O objetivo da figura é meramente de esclarecer a metodologia utilizada.

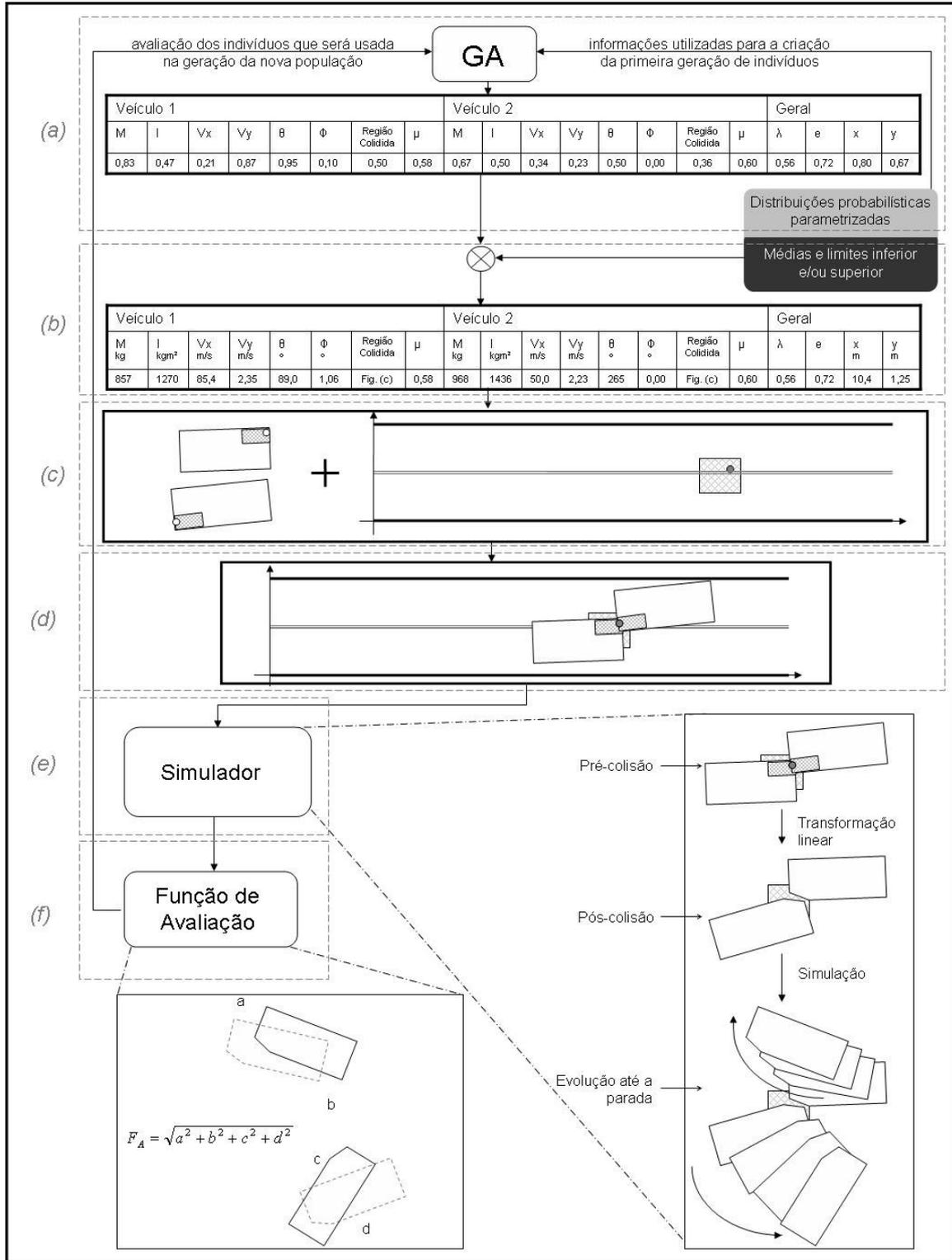


Figura 22 - Detalhamento do GA conjugado ao programa de simulação

Para não haver dificuldades de interpretação das velocidades e atitudes dos veículos por parte do usuário, estas serão definidas nas coordenadas locais. As transformações para coordenadas globais, serão efetuadas no próprio programa de simulação.

Resolvidos os desafios encontrados, tem-se como padrão de solução as seguintes etapas:

1. O usuário entra com os limites inferior e superior das variáveis a serem tratadas. Em geral as variáveis serão, como citado acima, as velocidades iniciais, posições iniciais, atitudes iniciais, partes colididas dos veículos, local da colisão. Cada valor atribuído a cada uma destas variáveis será, por definição, um gene do algoritmo. Cada conjunto de valores, sendo cada um deles atribuído a uma das variáveis, contemplando todas as variáveis, será portanto um indivíduo ou cromossomo. O usuário deve fornecer também neste momento, os parâmetros constantes da colisão, como as características do carro (bitolas, distâncias entre eixos, massa,...) e características de interação (atrito pneu-solo, coeficiente de interpenetração,...).
2. O usuário entra com as posições-alvo finais, ou seja, aquelas conhecidas da cena do acidente, também denominadas posições reais.
3. O algoritmo genético irá então “criar” a primeira população através de geração aleatória de valores para as variáveis, respeitando obviamente os limites impostos pelo usuário na etapa 1. A população terá o tamanho, ou o número de indivíduos estipulados no programa, podendo o usuário mudar este tamanho ou número a qualquer momento, caso haja necessidade.
4. Cada um destes indivíduos contém todas as informações necessárias para uma simulação completa. Com isso uma cena do acidente poderá ser montada, ou seja, através da definição do ângulo de cada veículo, das partes colididas e do local da colisão, os veículos serão colocados em posição de colisão. Esta será então simulada e as posições finais dos veículos serão comparadas com aquela fornecida pelo usuário na etapa 2. Para esta comparação será utilizada o que se denomina função de avaliação, definida anteriormente, que tem por objetivo quantificar quanto uma configuração está próxima da real.
5. Através de cada função de avaliação, cada indivíduo terá suas chances de se reproduzir, ou seja, de misturar seus valores com o de outro indivíduo. Para tanto, os indivíduos com melhores funções de avaliação, que neste caso nada mais são que funções de menores magnitudes (mais próximas de

zero), terão maiores chances de se reproduzirem. O critério específico utilizado para definir qual a diferença de probabilidade de escolha entre os indivíduos mais aptos e aqueles menos aptos será dado pela função de seleção. As misturas destes valores são feitas gene a gene, ou seja, velocidade inicial de um indivíduo com velocidade inicial de outro indivíduo, local de colisão de um indivíduo com local de colisão de outro indivíduo e assim por diante. As misturas também respeitam os limites impostos pelos dois indivíduos e consequentemente os limites impostos inicialmente. Por exemplo, se um indivíduo tem sua velocidade inicial estipulada em 10 m/s e vai se misturar com um indivíduo que tem a velocidade inicial de 12 m/s, os dois indivíduos resultantes ficarão com suas velocidades entre estes dois valores respeitando a seguinte regra:

$$\text{Velocidade do primeiro "filho"} = 10 + (1-p) \cdot (12-10)$$

$$\text{Velocidade do segundo "filho"} = 10 + p \cdot (12-10)$$

onde p é um número aleatório entre 0 e 1 gerado GA.

6. Outra forma utilizada para criar indivíduos para a nova geração é a mutação. Esta escolhe aleatoriamente, e normalmente a taxas muito baixas (em torno de 5%), indivíduos que não se cruzarão com nenhum outro, mas terão somente algumas de suas características alteradas aleatoriamente (dentro dos limites iniciais) e serão colocados desta forma na nova geração. Estes indivíduos são responsáveis pela procura de ótimos em outras regiões, caso a região de ótimo procurada pela maioria seja local e não global. As chances de se ter mutação na população vão decaindo com o tempo, já que a pesquisa de outras regiões se mostra interessante somente no começo da busca.
7. A última forma utilizada na criação de novas gerações é a perpetuação dos melhores indivíduos. A quantidade de indivíduos mantidos também pode ser alterada pelo usuário e garante que as melhores combinações de genes (valores) até um dado momento não serão perdidas na mistura com outros cromossomos (indivíduos).
8. Através dos três pontos anteriores tem-se a nova população ou geração. Com esta nova população retorna-se à etapa 4 e se recomeça o processo para uma nova geração. Assim sucessivamente até que algum critério de parada seja alcançado, a saber:
 - Limite inferior da função de avaliação – este é o principal critério e a partir deste objetivo os parâmetros do GA são configurados. Se a função de avaliação do melhor indivíduo estiver abaixo do limite

estipulado pelo usuário, o otimizador pára e toma este indivíduo como resultado final. Em particular, o valor da função de avaliação da forma que se encontra em sua versão final utilizada neste trabalho retorna o limite superior para as quatro distâncias medidas e a metade deste valor é a média destas distâncias.

- Número máximo de gerações, ou seja, uma vez alcançado o número estipulado, o otimizador pára e assume como resultado final o melhor conjunto de valores até então, ou seja, aquele conjunto que tem a menor função de avaliação.
- Limite máximo de tempo em estagnação, ou seja, a quantidade de segundos que o programa ficará calculando sem achar um indivíduo melhor que aquele dito como o atual melhor.
- Limite máximo de gerações em estagnação. Análogo ao anterior, somente tendo como critério o número de gerações, e não o tempo. Estes dois últimos são equivalentes e portanto não faz sentido a utilização simultânea de ambos.

Na Figura 24 encontra-se o fluxograma relativo ao procedimento de otimização como um todo. Neste fluxograma não se mostra os dois últimos critérios de parada citados anteriormente, já que estes não são habitualmente utilizados.

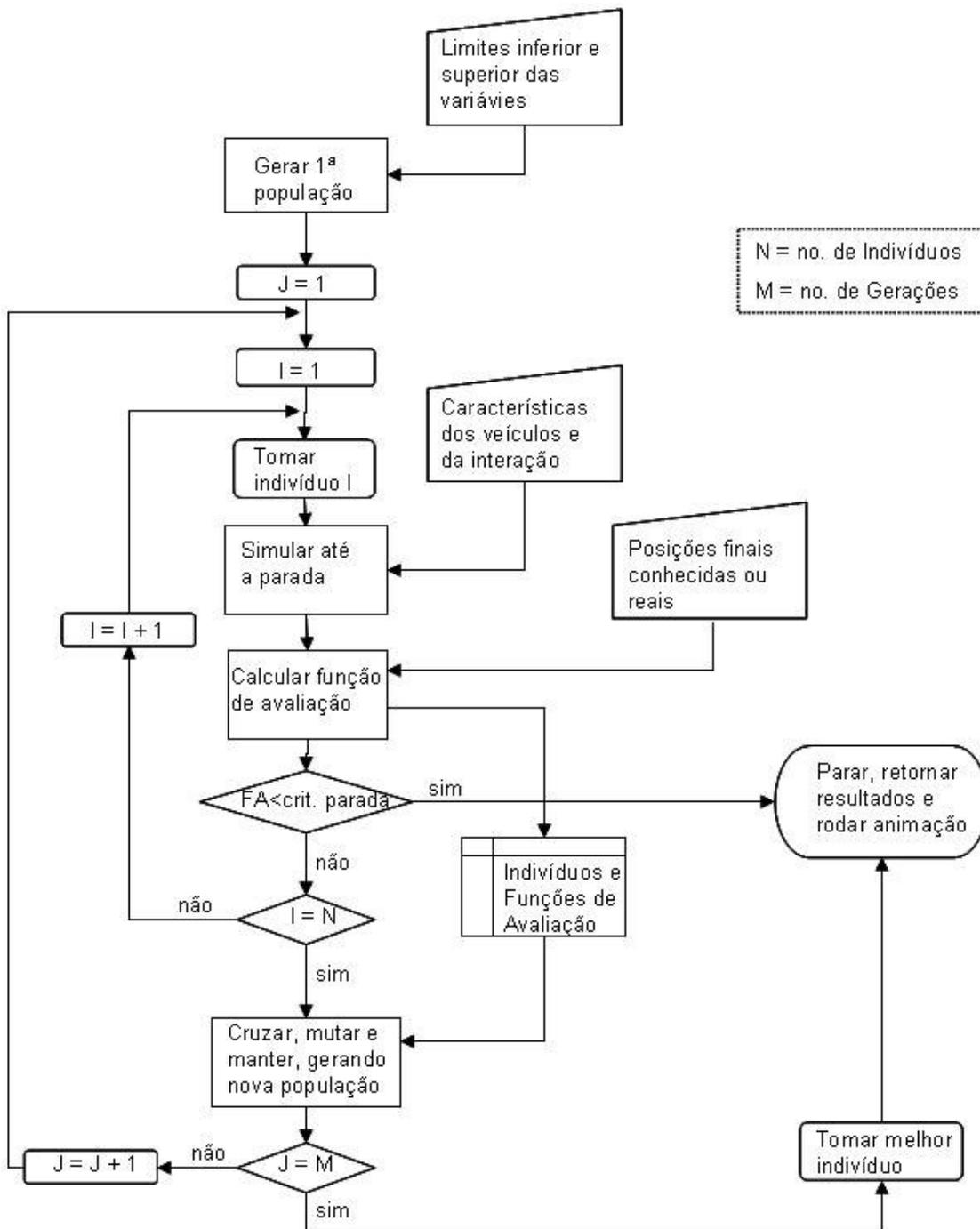


Figura 23 – Fluxograma do Procedimento de Otimização