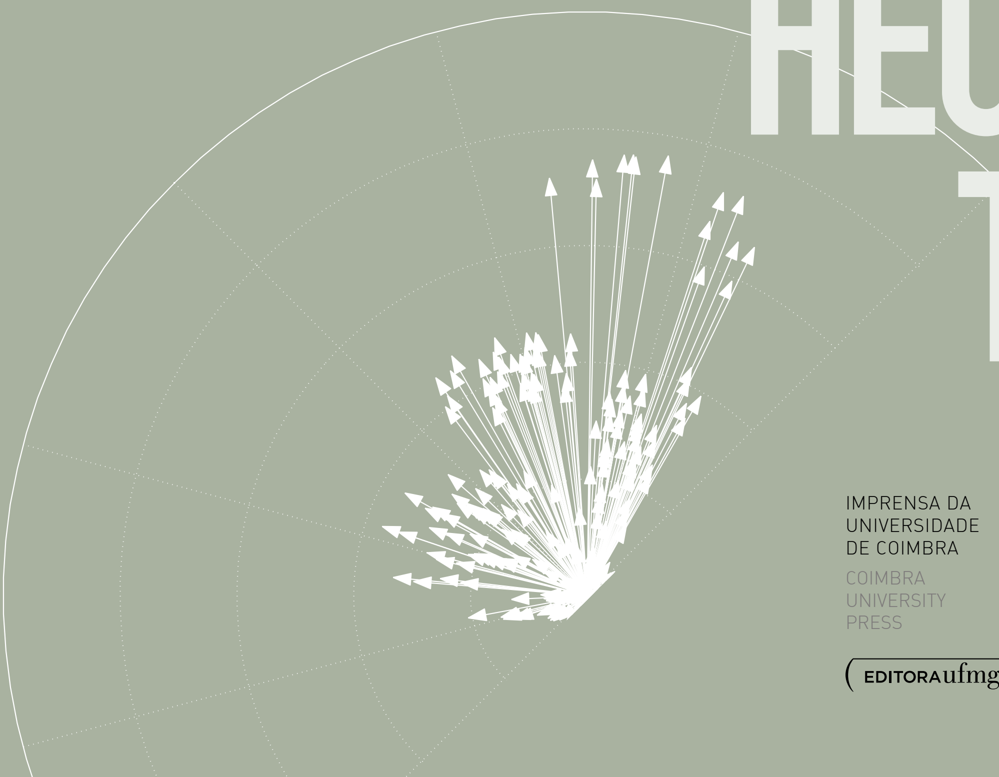
# MANUAL DE COMPUTAÇÃO EVOLUTIVA E METAHEURÍSTICA

ANTÓNIO GASPAR-CUNHA
RICARDO TAKAHASHI
CARLOS HENGGELER ANTUNES
COORDENADORES



IMPRENSA DA UNIVERSIDADE DE COIMBRA
COIMBRA UNIVERSITY PRESS

EDITORA UFMG

## CAPÍTULO 14

# Tratamento de Restrições em Algoritmos Evolutivos

*Elizabeth F. Wanner*

*Departamento de Computação*
*Centro Federal de Educação Tecnológica de Minas Gerais*

Os algoritmos evolutivos (AEs) são classificados como métodos de otimização global e são considerados métodos robustos e efetivos quando se deseja encontrar um mínimo global aproximado. A natureza exploratória dos AEs permite uma identificação rápida de regiões promissoras no espaço de busca, mas, por outro lado, prejudica a convergência e precisão nos estágios finais do processo de busca. Nesses estágios finais pouca informação nova é obtida através de seus mecanismos de busca local, enquanto seus mecanismos de busca global introduzem perturbações que são muito grandes para permitir uma convergênca com alta precisão ((De Jong, 1993; Mitchell, 1996)). Os AEs não utilizam informações locais do espaço de soluções, tais como derivadas, durante o processo de busca, o que faz com que esses algoritmos apresentem uma taxa de convergência mais lenta se comparados às técnicas clássicas de busca local (veja (Jonhson e Rahmat-Samii, 1997)). Além disso, os AEs, por serem algoritmos irrestritos por natureza, possuem dificuldade de localizar o ponto de ótimo em problemas com restrições, principalmente de igualdade. Esta dificuldade é devida à inabilidade do AE em realizar uma busca em regiões de volume zero, regiões que possuem dimensão menor que o espaço de busca original.

Os AES, originalmente criados para lidar com problemas irrestritos, necessitam de mecanismos específicos para incorporar restrições à função objetivo. Inicialmente em (Michalewicz, 1995a) e posteriormente em (Michalewicz e Schoenauer, 1996a), os autores discutiram os diversos métodos de tratamento de restrições em uma classe especial dos AES, os algoritmos genéticos. A maioria das técnicas de tratamento de restrições foi classificada em cinco categorias:

1. Métodos baseados em funções de penalidade;
2. Métodos baseados na preservação da factibilidade das soluções;
3. Métodos baseados na distinção entre soluções factíveis e infactíveis;
4. Métodos baseados em representações e operadores;
5. Métodos híbridos.

A primeira classe de métodos utiliza vários tipos de funções de penalidade para penalizar soluções que violem alguma restrição. A segunda classe de métodos usa explicitamente o conhecimento da estrutura das restrições e utiliza vários operadores de busca para manter a factibilidade das soluções. A terceira classe de métodos utiliza operadores de busca distintos para lidar com as soluções factíveis e infactíveis, enquanto a quarta classe usa um esquema de representação indireta que contêm instruções de como construir uma solução factível. Finalmente, na quinta classe, os AEs são combinados com outras heurísticas ou com outros métodos clássicos de tratamento de restrições.

Neste capítulo, algumas técnicas especializadas no tratamento de restrições, baseadas nessa classificação, serão apresentadas e discutidas. Os métodos baseados em otimização multiobjetivo também serão abordados. Para obter maiores detalhes destas e de outras técnicas especializadas no tratamento de restrição, recomenda-se a leitura dos trabalhos (Michalewicz, 1995b; Coello Coello, 2002; Venkatraman e Yen, 2005).

Em todos os métodos que serão discutidos nesse capítulo, considere o seguinte problema de otimização mono-objetivo:

$$\begin{array}{l} x^* = \min_x f(x) \\ \text{sujeito a:} \begin{cases} g_i(x) \leq 0; & i = 1, 2, \cdots, r \\ h_j(x) = 0; & j = 1, 2, \cdots, p \end{cases} \end{array} \tag{14.1}$$

sendo que $x \in \mathbb{R}^n$, $f(\cdot) : \mathbb{R}^n \to \mathbb{R}$, $g(\cdot) : \mathbb{R}^n \to \mathbb{R}^r$, e $h(\cdot) : \mathbb{R}^n \to \mathbb{R}^p$. As funções $g_i$ e $h_j$ representam, respectivamente, funções de restrição de desigualdade e de igualdade.

Usualmente, cada restrição de igualdade do problema (14.1) deve ser relaxada e transformada em duas restrições de desigualdade da forma:

$$\begin{aligned} h_j(x) - \epsilon \leq 0 \\ -h_j(x) - \epsilon \leq 0 \end{aligned}$$

onde $\epsilon$ é um número pequeno. Desta forma, o problema (14.1) é transformado no seguinte problema mono-objetivo com apenas restrições de desigualdade:

$$\begin{array}{l} x^* = \min_x f(x) \\ \text{sujeito a:} \begin{cases} g_i(x) \leq 0; & i = 1, 2, \cdots, r \\ g_k(x) = h_j(x) - \epsilon \leq 0 & k = r+1, \cdots, r+p \\ g_j(x) = -h_j(x) - \epsilon \leq 0 & j = r+p+1, \cdots, r+2p \end{cases} \end{array} \tag{14.2}$$

Observe que o problema considerado é o de minimizar uma função objetivo. Entretanto, a maximização de uma função objetivo pode ser convertida em um problema de minimização através do princípio de dualidade, no qual

$$\max\ f(x) = \min\ - f(x)$$

.

Este capítulo está dividido da seguinte forma:

**Seção 1** descreve a maneira usual de tratamento de restrições, o método de penalidade, e algumas variações deste;

**Seção 2** descreve os métodos que se baseiam na premissa de que soluções factíveis são preferíveis à soluções infactíveis;

**Seção 3** apresenta alguns métodos que fazem uso de representações e operadores especiais, responsáveis pela preservação da factibilidade das soluções;

**Seção 4** descreve os métodos que transformam o problema mono-objetivo em um problema multiobjetivo ao tratar as restrições como funções objetivo adicionais;

**Seção 5** apresenta metodologias que utilizam alguma técnica de otimização, geralmente numérica ou heurística, combinada com um algoritmo evolutivo;

**Seção 6** descreve algumas técnicas de tratamento de restrições em algoritmos multiobjetivo;

**Seção 7** conclui o capítulo.

## 1. Métodos de Penalidade

A maneira tradicional de incorporar restrições nos algoritmos evolutivos é através do método de penalidade, que transforma o problema restrito (14.1) em um problema irrestrito aproximadamente equivalente. No método de penalidade, um termo é adicionado à função objetivo penalizando qualquer violação das restrições. Este método gera uma sequência de pontos infactíveis cujo limite corresponde à solução ótima do problema original.

As restrições de desigualdade e de igualdade são incorporadas à função de aptidão $F(x)$, que passa ser definida como a soma da função-objetivo $f(x)$ e o termo de penalidade, o qual depende da violação das restrições de desigualdade $g_i(x)$ e de igualdade $h_j(x)$:

$$F(x) = f(x) + \sum_{i=1}^{r} \alpha_i [\max\{0, g_i(x)\}]^{\gamma} + \sum_{j=1}^{p} \alpha_j |h_j(x)|^{p} \tag{14.3}$$

onde, usualmente $\gamma$ e $p$ são iguais a 1 ou 2 O parâmetro $\alpha_i$ corresponde ao parâmetro de penalidade para a $i$-ésima restrição de desigualdade e o parâmetro $\alpha_j$ corresponde ao parâmetro de penalidade para a $j$-ésima restrição de igualdade.

Ao se utilizar o método de penalidade, o valor total das violações das restrições é usado para penalizar uma solução infactível. Assim, soluções factíveis são favorecidas durante o processo de seleção. Apesar da popularidade, o método de penalidade apresenta alguns aspectos negativos, como por exemplo a falta de regras bem definidas para se determinar os parâmetros de penalidade. Esses parâmetros são responsáveis pelo grau de penalização a ser aplicado e, para guiar a busca de modo eficiente para a região factível, devem ser bem escolhidos. Assim, existem $r + 2p$ parâmetros distintos que precisam ser escolhidos de modo eficiente nesta abordagem. Com o objetivo de se reduzir o número de parâmetros, as funções de restrições são normalizadas e apenas um valor do parâmetro $\alpha$ pode ser escolhido para todas as restrições do problema.

Independentemente do número de parâmetros envolvidos na formulação, a solução ótima para a função (14.3) depende da escolha desses parâmetros. A solução do problema penalizado pode estar arbitrariamente perto da solução original se valores grandes de $\alpha$ forem escolhidos.

Este fato pode ser observado no seguinte exemplo:

$$x^* = \min_x x_1^2 + x_2^2$$
$$\text{sujeito a:} \left\{ \; x_1 + x_2 - 1 = 0 \right.$$

A solução ótima deste problema é o ponto $x^* = (0.5; 0.5)$ e o valor ótimo é $f(x^*) = 0.5$. Considere agora o problema penalizado irrestrito:

$$x^* = \min_x x_1^2 + x_2^2 + \alpha(x_1 + x_2 - 1)^2$$

Observe que, para qualquer valor de $\alpha$, a função objetivo deste novo problema é convexa. Assim, a condição necessária e suficiente de otimalidade é dada por $\nabla(xx_1^2 + x_2^2 + \alpha(x_1 + x_2 - 1)^2) = 0$, e portanto temos que

$$x_1 + \alpha(x_1 + x_2 - 1) = 0$$
$$x_2 + \alpha(x_1 + x_2 - 1) = 0$$

Resolvendo-se essa duas equações simultaneamente obtemos

$$x_1 = x_2 = \frac{\alpha}{2\alpha + 1}$$

. Portanto, se $\alpha \to \infty$ temos que $x_i = x_2 \to 0.5$

Entretanto, escolhendo-se um valor muito grande para $\alpha$ a factibilidade será enfatizada e, a maior parte dos algoritmos irá convergir rapidamente para um ponto factível qualquer. E mesmo que esse ponto factível esteja longe da solução ótima, uma convergência prematura poderá ocorrer. Além disso, se a solução ótima estiver localizada na fronteira da região factível, o algoritmo será atraído para o interior da região factível e não será capaz de se mover para a fronteira entre as regiões factível e infactível. Portanto, com valores elevados para $\alpha$, o algoritmo não será capaz de explorar regiões infactíveis desde o início do processo de busca. E, se existirem regiões factíveis disjuntas, o algoritmo irá mover-se para uma dessas regiões e não será capaz de explorar as outras regiões factíveis, a menos que elas estejam muito próximas uma da outra.

Por outro lado, escolhendo-se um valor muito pequeno pequeno para $\alpha$, o algoritmo passará muito tempo explorando regiões infactíveis e convergirá para pontos fortemente infactíveis ((Bazaraa et al., 1979)). Este é um aspecto muito importante em AEs, já que vários problemas possuem a solução ótima na fronteira da região factível.

Existem pelo menos três escolhas que definem uma relação entre um indivíduo infactível e a região factível do espaço de busca ((Dasgupta e Michalewicz, 1997)):

1. um indivíduo deve ser penalizado apenas pela infactibilidade, sem levar-se em conta a proximidade do mesmo com a região factível;

2. o valor total das violações das restrições pode ser calculado e ser utilizado para definir um fator de penalidade;

3. o custo de tornar uma solução infactível em uma factível deve ser levado em consideração.

Com o objetivo de ajustar os parâmetros de penalidade ou mesmo de reduzir os efeitos desses, várias heurísticas para o desenvolvimento de novas funções de penalidade foram propostos. Estas heurísticas se baseiam nessas três definições dadas acima. Iremos, brevemente, fazer uma revisão de alguns desses métodos.

## Penalidades Estáticas

Nessa classe de funções de penalidade, os parâmetros de penalidade não dependem da geração atual do algoritmo evolutivo e, portanto, permanecem constantes durante todo o processo evolutivo.

Em (Homaifar et al., 1994), os autores propuseram uma metodologia na qual o usuário pré-definia níveis de violação das restrições. Para cada nível, o parâmetro de penalidade era escolhido de modo que este aumentava sempre que o nível aumentava. Considerando um problema de forma (14.2), a população inicial era composta de indivíduos factíveis e infactíveis, sendo a função de aptidão dada por:

$$F(x) = f(x) + \sum_{i=1}^{r+2p} R_{k,i} \times \max\left[0,\ g_i(x)^2\right]$$

sendo $R_{i,k}$ os parâmetros de penalidade, e $k = 1, 2, \cdots, l$ onde $l$ representava o número de níveis de violação das restrições pré-definido pelo usuário. Essa metodologia tinha como objetivo balancear cada restrição separadamente definindo um conjunto de parâmetros de penalidade para cada restrição. Esses parâmetros eram determinados através de certas regras determinísticas pré-definidas.

Uma desvantagem desta metodologia se deve ao número elevado de parâmetros envolvidos: com $r + 2p$ restrições e $l$ níveis de violação, são necessários $(r + 2p)(2l + 1)$ parâmetros. Além disso, é necessário se conhecer os graus de violação das restrições em cada problema, os quais não são conhecidos em qualquer problema.

## Penalidades Dinâmicas

Nessa classe de métodos, a geração na qual o AE se encontra influencia o parâmetro de penalidade que será aplicado. Usualmente, o parâmetro de penalidade aumenta à medida que o número de gerações aumenta.

Em (Joines e Houck, 1994), os autores propuseram uma metodologia na qual a função de aptidão, a cada geração t, era dada por:

$$F(x) = f(x) + (C \times t)^{\alpha} \times SVC(\beta, x)$$

sendo $C$, $\alpha$ e $\beta$ constantes definidas pelo usuário,

$$SVC(\beta, x) = \sum_{i=1}^{r} D_i^{\beta}(x) + \sum_{j=1}^{p} D_j(x)$$

e

$$D_i(x) = \begin{cases} 0 & g_i(x) \leq 0 \\ |g_i(x)| & g_i(x) > 0 \end{cases}$$

$$D_j(x) = \begin{cases} 0 & -\epsilon \leq h_j(x) \leq \epsilon \\ |h_j(x)| & \text{nos outros casos} \end{cases}$$

Alguns autores acreditam que os métodos de penalidade dinâmica funcionam melhor que os métodos de penalidade estática. Entretanto, na prática, é tão difícil desenvolver um método de penalidade dinâmica quanto encontrar parâmetros de penalidade que funcionem bem em métodos de penalidade estática.

Por exemplo, neste método descito acima verificou-se que a qualidade da solução final obtida era extremamente sensível à escolha dos parâmetros $\alpha$ e $\beta$. Além disso, ou o método convergia para soluções infactíveis, ou para uma solução factível longe da solução ótima ((Dasgupta e Michalewicz, 1997)).

### Penalidades de Recozimento

Em (Michalewicz e Attia, 1994), os autores desenvolveram um método baseado na idéia do algoritmo de Recozimento Simulado ((Kirkpatrick Jr. et al., 1983)) de forma que os parâmetros de penalidade eram modificados apenas quando o algoritmo ficasse preso em um ponto de mínimo local. Apenas as restrições ativas eram consideradas a cada iteração e o parâmetro de penalidade aumentava ao longo das gerações, de modo que as soluções infactíveis eram bastante penalizadas na última geração. As restrições eram divididas em quatro grupos: restrições de igualdade lineares, restrições de desigualdade lineares, restrições de igualdade não lineares e restrições de desigualdade não lineares. Um conjunto de restrições ativa $A$ era construído e todas as restrições não lineares de igualdade, juntamente com todas as restrições de desigualdade não lineares ativas, eram incluídas nesse conjunto.

A função de aptidão era dada por:

$$F(x) = f(x) + \frac{1}{2\tau} \sum_{i \in A} \phi_i^2(x)$$

sendo $\tau$ um valor de *temperatura* e

$$\phi_i(x) = \begin{cases} \max[o,\ g_i(x)] & 1 \leq i \leq r \\ |h_j(x)| & 1 \leq j \leq p \end{cases}$$

Um aspecto interessante dessa técnica é que a população inicial era composta de múltiplas cópias de uma única solução satisfazendo todas as restrições. A cada geração, a temperatura $\tau$ diminuia e uma nova população era criada usando a melhor solução obtida na geração anterior. Essa solução era usada como ponto inicial para gerar as cópias para a próxima população. O processo terminava quando uma temperatura pré-definida $\tau_f$ era atingida.

Este algoritmo era capaz de preservar a factibilidade de todas as restrições lineares através de operadores que levavam indivíduos factíveis em indivíduos factíveis. A cada geração o algoritmo considerava apenas as restrições ativas, o que fazia com que a pressão exercida sobre as soluções infactíveis diminuísse devido ao decréscimo da temperatura $\tau$. Este método era bastante sensível a escolha dos valores de $\tau$, problema similar ao que occore no algoritmo de Recozimento Simulado. Essa metodologia de lidar separadamente com as restrições lineares pode ser considerada eficiente, porém necessita de uma solução inicial factível e de operadores especiais que preservem a factibilidade.

### Penalidades Adaptativas

Aqui os parâmetros de penalidade são alterados utilizando informações adquiridas durante o processo evolutivo.

Este método foi inicialmente proposto em (Bean e Alouane, 1992) e a função de aptidão era dada por:

$$F(x) = f(x) + \lambda(t) \left[ \sum_{i=1}^{r} g_i^2(x) + \sum_{j=1}^{p} |h_j(x)| \right]$$

O parâmetro $\lambda(t)$ era atualizado a cada geração $t$, de acordo com a seguinte regra

$$\lambda(t+1) = \begin{cases} \left(\frac{1}{\beta_1}\right)\lambda(t) & b_i \in \mathcal{F} \quad \forall \quad t-g+1 \leq i \leq t \\ \\ \beta_2\lambda(t) & b_i \notin \mathcal{F} \quad \forall \quad t-g+1 \leq i \leq t \\ \\ \lambda(t) & \text{nos outros casos} \end{cases}$$

sendo $b_i$ o melhor indivíduo da geração $i$, $\mathcal{F}$ o conjunto factível, $\beta_1 \neq \beta_2$ e $\beta_1,\ \beta_2 > 1$.

Neste método, o parâmetro de penalidade da próxima geração $\lambda(t+1)$ diminuia quando todos os melhores elementos nas últimas $g$ gerações fossem factíveis, aumentava se todos os melhores elementos fossem infactíveis e permanecia o mesmo da geração anterior nos outros casos. Os valores de $g$, $\beta_1$ e $\beta_2$ precisam ser bem escolhidos para um bom resultado do método.

Em (Lemonge e Barbosa, 2004), os autores propuseram um método de penalidade adaptativa sem a necessidade de parâmetros externos. A adaptação utilizava informações da população corrente, tais como a média dos valores das funções objetivo e da violação de cada restrição. A função de aptidão era dada por:

$$F(x) = \begin{cases} f(x) & \text{se x for factível} \\ \\ \bar{f}(x) + \displaystyle\sum_{i=1}^{r+2p} k_i v_i(x) & \text{se x não for factível} \end{cases}$$

sendo que

$$\bar{f}(x) = \begin{cases} f(x) & \text{se } f(x) > \langle f(x) \rangle \\ \\ \langle f(x) \rangle & \text{se } f(x) \leq \langle f(x) \rangle \end{cases}$$

O valor $\langle f(x) \rangle$ representava a média dos valores das funções objetivo na população corrente. O parâmetro de penalidade $k_i$ era definido a cada geração por

$$k_i = |\langle f(x) \rangle| \frac{\langle v_i(x) \rangle}{\displaystyle\sum_{l=1}^{r+2p} \langle v_l(x) \rangle^2}$$

sendo $\langle v_l(x) \rangle$ a média da violação da $l$-ésima restrição na população corrente. A idéia básica era que os valores dos parâmetros de penalidade fossem distribuídos de modo que aquelas restrições que eram mais difícieis de serem atingidas teriam um valor mais alto para o parâmetro de penalidade.

### Penalidade por Morte

O método de penalidade por morte é possivelmente o método mais simples de tratar restrições. Neste método, os indivíduos infactíveis são simplesmente rejeitados. No caso de um indivíduo infactível, a função de aptidão recebe o valor zero, não sendo mais necessário o cálculo do grau de violação das restrições.

Este método funciona razoavelmente bem quando a região factível é convexa e representa uma parte considerável da região de busca total[1]. Nas outras situações este método apresenta algumas limitações. Em problemas onde a razão entre a região factível e a região de busca é pequena, e quando a população inicial só contenha soluções infactíveis, é necessário que as soluções possam ser melhoradas e não simplesmente eliminadas. Além disso, é frequente situações nas quais o algoritmo chega ao ótimo do problema mais facilmente quando este consegue cruzar a fronteira entre as regiões infactível e factível, especialmente quando a região factível for não convexa, não sendo portanto aconselhável uma eliminação de indivíduos infactíveis.

[1] Define-se a *região de busca* como a parcela do espaço de busca cujos pontos são representados pela codificação empregada no algoritmo evolutivo.

## 2. Métodos de Preservação da Factibilidade

Nesta classe de métodos, o algoritmo evolutivo utiliza, explicitamente, o conhecimento da estrutura das restrições e utiliza vários operadores de busca para manter a factibilidade das soluções. Esses métodos se baseiam na premissa de que soluções factíveis são preferíveis à soluções infactíveis.

O método proposto em (Michalewicz e Janikow, 1991) considerava o problema mono-objetivo (14.1), com apenas restrições lineares, e iniciava-se com uma solução factivel ou com uma população de soluções factíveis. Os operadores genéticos eram responsáveis pela preservação da factibilidade das novas soluções. No caso da mutação, quando uma determinada componente $x_i$ do vetor $x$ era mutada, o domínio desta componente era calculado. Este domínio era determinado levando-se em consideração as restrições lineares e as outras componente do vetor $x$. Assim, o novo valor de $x_i$ era escolhido neste domínio determinado e, portanto, o novo vetor $x$ seria sempre factível. O operador de cruzamento entre duas soluções $x$ e $y$ era dado por

$$ax + (1 - a)y \quad 0 \leq a \leq 1$$

Observe que, como o problema só possuia restrições lineares, era possível garantir a convexidade da região de busca e, portanto, esse operador de cruzamento sempre gerava soluções factíveis.

Este método não necessita de nenhum parâmetro adicional além dos parâmetros usuais dos AEs e pode ser generalizado para lidar com restrições não lineares desde que a região de busca seja convexa. Entretanto, não é capaz de lidar com regiões de busca não convexas, que representam a situação geral em problemas com restrições não lineares.

O método proposto em (Powell e Skolnick, 1989) incorporava uma regra heurística sugerida em (Richardson et al., 1993) que sugeria um mapeamento de soluções factíveis no intervalo $(-\infty,\ 1)$ e de soluções infactíveis no intervalo $(1,\ \infty)$, considerando sempre as soluções factíveis melhores que as infactíveis. Considerando o problema mono-objetivo (14.1), os indivíduos da população eram avaliados usando a seguinte função de aptidão:

$$F(x) = \begin{cases} f(x) & \text{se factível} \\ 1 + r\left(\displaystyle\sum_{i=1}^{r} g_i(x) + \sum_{j=1}^{p} h_j(x)\right) & \text{se infactível} \end{cases}$$

A função $f(x)$ era normalizada no intervalo $(-\infty,\ 1)$, as funções $g_i(x)$ e $h_j(x)$ eram normalizadas no intervalo $(1,\ \infty)$ e $r$ era uma constante real. Observe que, nesse método, a função objetivo e o valor total das violações das restrições não eram combinados quando o indivíduo era infactível, como ocorre nos métodos de penalidade. Soluções infactíveis era comparadas baseando-se apenas na violação das restrições, enquanto que soluções factíveis eram comparadas baseando-se apenas no valor da função objetivo.

Este método apresentava algumas características importantes: enquanto nenhuma solução factível fosse obtida, o valor da função objetivo não influenciaria na função de aptidão; tendo soluções factíveis e infactíveis na população as soluções factíveis teriam um valor melhor da função de aptidão. A desvantagem desse método deve-se à falta de diversidade no mecanismo de seleção. Com o objetivo de evitar uma convergência prematura, os autores usaram um processo de seleção com classificação linear, de modo que nas gerações iniciais a convergência fosse lenta e nas gerações finais a convergência fosse forçada devido ao número de cópias presentes do melhor indivíduo.

O método proposto em (Deb, 2000) utilizava um operador de seleção, baseado em torneio, no qual duas soluções eram comparadas a cada vez, seguindo os seguintes critérios:

1. Qualquer solução factível é preferida em relação à uma solução infactível.

2. Entre duas soluções factíveis, aquela que tiver melhor valor de função objetivo será preferida.

3. Entre duas soluções infactíveis, aquela que tiver menor violação das restrições será preferida.

Usualmente os parâmetros de penalidade são necessários para definir uma violação nas restrições relacionando-as ao valor da função objetivo. Entretanto, nesse método não existia a necessidade de definir um fator de penalidade, uma vez que nas três situações descritas acima as soluções nunca eram comparadas em termos dos valores da função objetivo e das restrições simultaneamente. Na primeira situação, nem o valor da função objetivo e nem o valor das restrições eram utilizados, uma vez que uma solução factível era sempre preferida. No segundo caso, soluções eram comparadas usando apenas os valores das funções objetivo, enquanto que no terceiro caso, as soluções eram comparadas usando apenas os valores das restrições violadas.

Considerando o problema mono-objetivo (14.2) e motivado por essas observações, a seguinte função de aptidão foi proposta, na qual as soluções infactíveis eram comparadas apenas em termos das restrições violadas:

$$F(x) = \begin{cases} f(x) & \text{se } g_i(x) \leq 0, \quad i = 1, \cdots, r+2p \\ f_{pior} + \sum_{i=1}^{r+2p} g_i(x) & \text{se } g_i(x) > 0, \quad i = 1, \cdots, r+2p \end{cases}$$

sendo que $f_{pior}$ correspondia ao valor da pior solução factível na população. Se na população não houvesse nenhuma solução factível então $f_{pior} = 0$. Observe que o valor de aptidão de uma solução infactível dependia do valor da violação das restrições e da população de soluções. Entretanto, o valor de aptidão de uma solução factível era fixo e igual ao valor da função objetivo.

Os resultados obtidos e apresentados em (Deb, 2000) podem ser considerados promissores, porém fica evidente que o método tem dificuldade em manter a diversidade na população. Para evitar uma convergência prematura, algumas técnicas de nicho combinadas com taxas maiores de mutação se fazem necessárias. Isso significa que o processo de busca foi direcionado inicialmente para procurar soluções factíveis e depois utilizou técnicas que mantinham a diversidade para se aproximar da solução ótima. As funções de partilha, envolvidas nas técnicas de nicho, são computacionalmente caras e introduzem um parâmetro extra ao algoritmo.

O conceito por trás destes dois métodos se baseia na hipótese de que soluções factíveis são melhores que soluções infactíveis, e supondo que esta hipótese seja válida, espera-se que esta técnica funcione bem. A diferença entre os dois métodos é que o segundo não necessita de um parâmetro de penalidade $r$. Entretanto estes métodos terão dificuldade em problemas nos quais a solução ótima está localizada na fronteira entre as regiões factível e infactível.

Em (Mezura-Montes e Coello Coello, 2005), os autores utilizaram um mecanismo de mutação auto-adaptativa em uma estratégia evolutiva do tipo $(\mu + \lambda)$ para explorar efetivamente a região de busca do problema restrito. Este algoritmo foi acoplado a um mecanismo de comparação baseado nas três regras de factibilidade descritas em (Deb, 2000) e a um mecanismo simples de diversidade. A idéia era permitir que indivíduos com baixo valor de violação de restrições e com o melhor valor de função objetivo fossem selecionados para a próxima geração. Este método, inicialmente, pretendia alcançar a região factível e posteriormente, pretendia mover pontos dentro da região factível na direção do ótimo do problema. Este método apresentou bons resultados em diversos problemas analíticos, sendo superior a alguns métodos de tratamento de restrições presentes na literatura.

## 3. Métodos Baseados em Representações e Decodificadores

Alguns tipos de representações especiais têm sido propostas para lidar com certos problemas nos quais uma representação genérica, por exemplo a representação binária usada tradicionalmente, não é apropriada. Devido à mudança na representação, é necessário que os novos operadores genéticos desenvolvidos funcionem de modo similar ao operadores genéticos usados na representação tradicional.

A mudança na representação tem como objetivo simplificar a forma da região de busca e os novos operadores genéticos são usualmente utilizados para preservar a factibilidade das soluções. A principal aplicação desses métodos é em problemas nos quais a localização de uma solução factível seja extremamente difícil. A referência (Davis, 1991) apresenta vários exemplos de AEs que utilizam representações e operadores especiais para resolver problemas reais que vão desde a problemas de horários, síntese de arquitetura de redes neurais a análise de DNA.

Um exemplo desse tipo de método é conhecido como *GEnetic algorithm for Numerical Optimization for COnstrained Problems* (GENOCOP). GENOCOP, proposto por (Michalewicz, 1992), eliminava as restrições de igualdade removendo parte da região de busca, e simplificando o problema para o AE. As restrições restantes eram restrições lineares de desigualdade, que formavam um conjunto convexo de busca para o algoritmo. O método buscava encontrar, inicialmente, uma solução factível através de uma amostragem da região factível. Se após um certo número de tentativas o método falhava em encontrar uma solução factível, o usuário devia escolher um ponto inicial aleátorio. A população inicial do algoritmo era composta por cópias dessa solução inicial. Os operadores genéticos de recombinação usavam combinação linear entre os indivíduos garantindo, através das propriedades de conjuntos convexos, a factibilidade dos descendentes.

Uma vez que o GENOCOP iniciava-se com um indivíduo factível, era necessário que o algoritmo (ou mesmo o usuário) fosse capaz de gerar uma solução factível em um tempo razoável. O fato de permitir apenas restrições lineares, limita a aplicação do método a regiões de busca convexas.

Uma outra alternativa pertencente a essa classe de métodos é a localização da fronteira da região factível. A idéia principal consistia em realizar a busca em áreas próximas à fronteira da região factível. Uma vez que problemas com restrições não-lineares possuem pelo menos uma restrição ativa na solução ótima, é justificável realizar uma busca na fronteira entre as regiões factível e infactível. Os componentes básicos desses tipos de métodos são

(a) mecanismo de geração de soluções factíveis iniciais e

(b) operadores genéticos especiais que sejam capazes de explorar a região factível.

Além disso, os operadores genéticos devem satisfazer mais duas condições ((Michalewicz, 1992)): os operadores de cruzamento precisam ser capazes de gerar todos os filhos entre os pais e pequenas mutações devem resultar em pequenas mudanças nos valor da função de aptidão.

Em (Michalewicz e Schoenauer, 1996b), operadores genéticos especializados foram desenvolvidos e baseados nas expressões das restrições. Supondo que a superfície de busca fosse uma superfície Riemaniana de dimensão $n-1$ no espaço $\mathbb{R}^n$ e regular [2], a distância euclidiana usual poderia ser usada como uma medida de distância nesta superfície. Os operadores, que levam pontos sobre a superfície em pontos sobre a superfície, foram baseados em curvas geodésicas (curvas geradas pela interseção da superfície com planos).

Observe que, nesta abordagem, os operadores eram dependentes da parametrização escolhida, fazendo que os operadores ficassem limitados a um certo tipo de problema. Entretanto nas situações em que os operadores foram aplicados, os resultados foram precisos e eficientes.

Nos métodos baseados em decodificadores, um cromossomo é responsável por fornecer instruções de como gerar uma solução factível. Cada decodificador impõe uma relação $T$ entre a solução factível e a solução decodificada. Algumas condições devem ser satisfeitas ao lidar com decodificadores ((Palmer e Kershenbaum, 1994b)):

1. para cada solução factível $s$ deve existir uma solução decodificada,

[2] uma superfície é dita regular quando o vetor gradiente, ortogonal à superfície, está definido em todos os pontos da superfície

2. cada solução decodificada $S$ deve corresponder a uma solução factível $s$,

3. todas as soluções factíveis devem ser representadas pelo mesmo número de decodificadores $d$,

4. a transformação $T$ deve ser rápida, e

5. a transformação $T$ deve possuir uma característica local de que pequenas mudanças na solução decodificada $S$ produzam pequenas mudanças na solução factível $s$.

Em (Koziel e Michalewicz, 1998, 1999), os autores propuseram um homomorfismo entre um cubo n-dimensional $[-1,\ 1]^n$ e uma região de busca, convexa ou não. A idéia principal consistia na transformação do problema mono-objetivo original (14.2) em um outro problema topologicamente equivalente, mas de fácil solução. Este mapeamento era responsável pela geração de soluções factíveis, não sendo necessária a aplicação de nenhum operador que mantivesse a factibilidade das soluções.

Esta metodologia mostrou-se eficiente porém alguns problemas ficaram evidentes: alto custo computacional e o homomorfismo $T$ não possuia a característica local de levar pequenas mudanças na solução decodificada em pequenas mudanças na solução factível quando a região de busca é disjunta.

## 4. Métodos Baseados em Otimização Multiobjetivo

A principal idéia por trás desses métodos resume-se em re-escrever o problema mono-objetivo restrito como um problema multiobjetivo. Em primeiro lugar, cada restrição de igualdade do problema (14.1) deve ser relaxada e transformada em duas restrições de desigualdade. Desta forma, o problema mono-objetivo passa a ter $r + 2p$ restrições de desigualdade e o novo problema multiobjetivo terá $r + 2p + 1$ funções objetivo. Qualquer técnica de otimização evolutionária multiobjetivo ((Fonseca e Fleming, 1995)) pode ser aplicada ao vetor $\vec{v} = (f(x),\ f_1(x),\ f_2(x),\ \cdots\ ,\ f_{r+2p}(x))$, no qual $f(x)$ corresponde a função objetivo original do problema (14.1) e $f_k(x),\quad k = 1, 2, \cdots, r + 2p$ representam as restrições do problema mono-objetivo. Para uma solução ser considerada ótima $x^*$, é necessário que $f_k(x^*) = 0$, $k = 1, 2, \cdots, r + 2p$ e $f(x^*) \leq f(y)$ para qualquer solução factível $y$.

As restrições em um problema mono-objetivo restrito podem ser vistas como objetivos que precisam ser fortemente atingidos antes que a função objetivo seja minimizada. Atender um conjunto de restrições violadas é claramente um problema multiobjetivo de minimizar novas funções até que um certo valor (meta) seja atingido. O conceito de não dominância é portanto aplicável.

Em (Coello Coello e Mezura-Montes, 2002), os autores usaram um algoritmo evolutivo multiobjetivo, com seleção baseada em torneio, e seguindo as regras básicas de factibilidade descritas em (Deb, 2000) e os critérios de dominância na população do algoritmo. Nesse método, um parâmetro externo $S_r$ foi usado para controlar a diversidade da população. Apesar de apresentar bons resultados nos problemas testados, o método apresentou dificuldades ao lidar com um aumento na dimensão do problema.

Uma heurística genérica para tratamento de restrições em algoritmos genético foi mostrada em (Venkatraman e Yen, 2005). A heurística proposta era dividida em duas fases. Na primeira fase, a função objetivo era desprezada e o problema de otimização restrito era tratado como um problema de satisfação das restrições. O algoritmo procurava minimizar a violação das restrições e, eventualente, encontrar uma solução factível. A solução com menor violação das restrições era arquivada e tratada como uma solução de elite na população. Na segunda fase, considerava-se um problema biobjetivo que minimizava a violação das restrições e o valor da função objetivo. Esta metodologia mostrou-se competitiva ao ser comparada com outras técnicas de tratamento de restrições.

Uma abordagem baseada em técnicas de otimização multiobjetivo foi proposta em (Camponogara e Talukdar, 1997). A metodologia consistia em transformar o problema mono-objetivo da forma (14.2) em um problema biobjetivo, minimizando a função objetivo e minimizando a seguinte função:

$$\phi(x) = \sum_{i=1}^{r+2p} \max\left[0,\ g_i(x)\right]$$

Redefinindo-se o problema como um problema biobjetivo, era possível gerar soluções não dominadas em relação aos dois objetivos. Cada duas soluções encontradas, $x_a$ e $x_b$, definiam uma direção de busca. Uma busca linear era feita nesta direção obtendo-se uma nova solução $x$ que dominava $x_a$ e $x_b$. Essa busca linear substituía o operador de cruzamento. Uma eliminação de metade da população era aplicada em intervalos regulares da seguinte forma: as soluções com menor valor da função de aptidão eram trocadas por soluções geradas aleatoriamente.

Esta metodologia apresenta problemas em manter a diversidade na população, uma vez que os piores indivíduos são descartados a cada geração. A utilização da busca linear provoca um aumento no custo computacional do algoritmo.

## 5. Métodos Híbridos

Nesta classe, são considerados as metodologias que utilizam alguma técnica de otimização, geralmente numérica ou heurística, para lidar com as restrições dentro de um algoritmo evolutivo.

Um algoritmo evolutivo híbrido que combina funções de penalidade com um método primal-dual foi proposto em (Adeli e Cheng, 1994). Este método considerava um problema mono-objetivo da forma (14.2) e baseava-se na minimização, em sequência, de métodos lagrangeanos, utilizando a seguinte função de aptidão:

$$F(x) = f(x) + \frac{1}{2}\sum_{i=1}^{r+2p} \gamma_i \left\{\left[g_i(x) + \mu_j\right]^+\right\}^2$$

sendo $\gamma_i > 0$, $\mu_i$ um parâmetro associado a $i$-ésima restrição e

$$\left[g_i(x) + \mu_j\right]^+ = \max\left[0, g_i(x) + \mu_j\right]$$

Os valores de $\mu_i$ eram definidos em termos da máxima violação da restrição correspondente. Um outro parâmetro adicional $\beta > 1$, definido pelo usuário, era multiplicado pelos valores antigos de $\mu_i$ e $\gamma_i$, fazendo com que a penalidade fosse aumentada com o aumento da geração. Apesar dos bons resultados apresentados por esse método em certos problemas, o método é dependente da escolhas destes parâmetros.

Em (Bernardino et al., 2009), um algoritmo genético (AG) foi hibridizado com um algoritmo baseado em Sistemas Imunes Artificiais (SIA) para lidar com restrições em problemas reais. O SIA, inspirado no príncípio de seleção clonal, foi acoplado ao AG para ajudar o algoritmo a guiar a população para a região factível. Este algoritmo híbrido foi testado em vários problemas com restrições de desigualdade e seu desempenho foi comparado com outras técnicas de tratamento de restrições, mostrando-se eficiente nos problemas testados.

Em (Araujo et al., 2009), um mecanismo de busca local foi acoplado a um algoritmo genético para resolver problemas mono-objetivo com restrições (lineares e não lineares) de desigualdade. Esse mecanismo de busca local utilizava aproximações quadráticas e lineares para as funções objetivo e de restrição. Na fase de busca local, estas aproximações definiam um problema associado no qual a função objetivo era quadrática e as funções de restrições eram quadráticas ou lineares. Esse problema associado era resolvido via uma formulação baseada em Desigualdades Lineares Matriciais ((Boyd et al., 1997)) e a solução deste problema associado era, então, introduzida na população do algoritmo genético. O algoritmo genético acoplado a esse mecanismo de busca local foi testado com vários problemas analíticos e os resultados mostraram que a metodologia é eficiente nesta classe de problemas.

Na maior parte dos trabalhos descritos nesse capítulo, as restrições de igualdade são relaxadas e transformadas em duas restrições de desigualdade. Técnicas apropriadas para tratamento de restrições de igualdade em algoritmos evolutivos não são frequentes. Restrições de igualdade constituem um problema para os AEs, uma vez que esse tipo de restrição define um conjunto factível de dimensão menor que a região de busca do problema.

Os AEs, heurísticas cuja busca é essencialmente baseada na amostragem de toda a região de busca, têm baixa probabilidade de encontrar soluções factíveis. Sabemos que a natureza aleatória dos operadores dos AEs permite a busca por regiões ótimas (bacias de atração) no espaço global. Esses operadores fazem com que a busca torne-se *espalhada* pelo espaço, maximizando a chance de encontrar outras bacias de atração. Podemos chamar essa propriedade de *busca por volume*. Entretanto, a aleatoriedade de tais operadores é conflitante com a necessidade de uma busca em objetos bem definidos de *volume zero*, tais como os conjuntos factíveis definidos por restrições de igualdade. A propriedade de *busca por volume* irá produzir um tipo de *movimento de afastamento aleatório* em relação ao objeto factível, o que acarretará na convergência a taxas mais lentas.

Tendo em mente essas dificuldades, uma metodologia especializada no tratamento de restrição de igualdade em algoritmos genéticos foi apresentada em (Wanner et al., 2005). Considerando um problema mono-objetivo com apenas uma restrição de igualdade, é possível construir uma aproximação quadrática para a função-objetivo e para a restrição, obtendo desta forma um problema quadrático associado. A solução analítica para esse problema associado pode ser obtida, porém, de uma maneira geral, essa solução não corresponde à solução do problema original. Entretanto, se as aproximações quadráticas forem precisas, a solução do problema quadrático será uma boa aproximação para a solução do problema original.

A idéia principal desse trabalho consistia em restringir o algoritmo a encontrar soluções que estivessem no interior de um objeto que possuia a dimensão igual à do conjunto factível. Este objeto consiste em uma aproximação de segunda ordem da superfície factível. Essa solução obtida deveria então ser introduzida na população do algoritmo. Desta forma, o operador de busca local proposto pode ser dividido em duas etapas:

- a aproximação de segunda ordem deve ser iterativamente atualizada, e
- o processo evolucionário de busca deve ser feito de forma que as soluções estejam na fronteira da superfície aproximada.

Um operador que segue essas etapas era, então, incluído no algoritmo genético e desta maneira, irá se opor ao *movimento de afastamento aleatório* descrito acima. Esse operador pode ser interpretado também como um operador de *elitismo* especializado em melhorar, iteração após iteração, a estimativa do conjunto factível. O operador de elitismo convencional, que mantém um único ponto (ou um conjunto de pontos) em cada geração, não promove uma busca pelo conjunto factível, uma vez que esse conjunto é uma superfície $m$-dimensional e não um conjunto discreto de pontos. O operador proposto garantia uma aproximação para esse objeto no seguinte sentido: a melhor superfície de segunda ordem que aproxima o conjunto factível era mantida e melhorada à medida que o algoritmo prosseguia.

Este operador pode ser descrito da seguinte forma:

**Passo 1.** Construir as aproximações quadráticas para a função-objetivo e para a restrição de igualdade

**Passo 2.** Construir o problema quadrático associado

**Passo 3.** Encontrar a solução analítica do problema quadrático associado

**Passo 4.** Introduzir essa solução na população atual do algoritmo genético

Este operador foi acoplado a um algoritmo genético e foi testado em problemas com restrições de igualdade. Em todos os casos, o algoritmo híbrido mostrou-se eficaz, encontrando a solução do problema original de modo preciso e eficiente.

Em (Peconick et al., 2007), os autores estenderam a metodologia descrita acima para tratar várias restriçôes não-lineares de igualdade com algoritmos genéticos. O novo operador era composto das seguintes etapas:

**Passo 1.** Construir as aproximações quadráticas para a função-objetivo e para as restrições de igualdade

**Passo 2.** Construir o problema quadrático associado

**Passo 3.** Determinar as projeções sobre as superfícies quadradas aproximadas via um processo iterativo, obtendo um ponto situado sobre todas as restrições

**Passo 3.** Introduzir este ponto na população atual do algoritmo genético

Esta metodologia mostrou-se eficaz nos problemas testados, atingindo a solução com mais precisão e eficiência. Observe que o operador proposto apenas resolve o problema de factibilidade. Ao obter soluções factíveis, o algoritmo genético irá guiar essas soluções na direção da solução ótima.

## 6. Tratamento de Restrições em Problemas Multiobjetivo

Considere o seguinte problema de otimização multiobjetivo:

$$x^* = \min_x \left\{ \begin{array}{l} f_i(x) \\ f_2(x) \\ \vdots \\ f_m(x) \end{array} \right. \tag{14.4}$$

$$\text{sujeito a:} \left\{ \begin{array}{ll} g_i(x) \leq 0; & i = 1, 2, \cdots, r \\ h_j(x) = 0; & j = 1, 2, \cdots, p \end{array} \right.$$

sendo que $x \in \mathbb{R}^n$, $f(\cdot) : \mathbb{R}^n \to \mathbb{R}^m$, $g(\cdot) : \mathbb{R}^n \to \mathbb{R}^r$, e $h(\cdot) : \mathbb{R}^n \to \mathbb{R}^p$. As funções $g_i$ e $h_j$ representam, respectivamente,as funções de restrição de desigualdade e de igualdade.

A maneira tradicional de tratar as restrições em problemas multiobjetivo é usar os métodos de penalidade descritos na seção 1. As restrições de desigualdade e de igualdade são incorporadas a $F_k(x)$, $k = 1, 2, \cdots, m$, e passa a ser definida como a soma da função-objetivo $f_k(x)$, $k = 1, 2, \cdots, m$ e do termo de penalidade, o qual depende da violação das restrições de desigualdade $g_i(x)$ e de igualdade $h_j(x)$:

$$F_k(x) = f_k(x) + \sum_{i=1}^{r} \alpha_i [\max\{0, g_i(x)\}]^\gamma + \sum_{j=1}^{p} \alpha_j |h_j(x)|^p \tag{14.5}$$

Em princípio, todas as estratégias de atualização dos parâmetros de penalidade descritas na seção 1 podem ser utilizados na formulação multiobjetivo. Entretanto, a maior parte dos trabalhos envolvendo problemas multiobjetivo com restrições utiliza a abordagem do método de penalidade estática ((Srinivas e Deb, 1994)). Uma vez que os termos de penalidade são adicionados a cada uma das funções objetivo, qualquer algoritmo evolutivo multiobjetivo pode ser utilizado. Espera-se que os algoritmos evolutivos sejam capazes de mover a população para o interior da região factível e encontrar boas estimativas para o conjunto Pareto-ótimo.

Em (Jiménez et al., 1999), os autores apresentaram um procedimento para tratamento de restrições de desigualdade em problemas multiobjetivo. Restrições de igualdade deviam ser relaxadas e transformadas em duas restrições de desigualdade. Este procedimento sugeria uma classificação cuidadosa das soluções factíveis e infactíveis e utilizava uma técnica de nicho para manter a diversidade. O algoritmo proposto utilizava um torneio binário baseado nas regras propostas em (Deb, 2000) e sugeria estratégias específicas para cada caso.

Este procedimento é mais sistemático do que a simples utilização de um método de penalidade, já que leva em consideração a factibilidade e a dominância das soluções. Entretanto, quando há um empate em relação à factibilidade ou à dominância, o algoritmo proposto tende a priorizar a dominância.

Em (Ray et al., 2000), uma técnica mais elaborada de tratamento de restrições foi proposta. Neste método, para construir uma nova população, eram realizadas três classificações baseadas nos critérios de não dominância: com relação aos valores das funções objetivo, com relação aos valores de violação de restrições, e com relação aos valores das funções objetivo e das violações das restrições. Uma vantagem deste método é a forma de preservar a diversidade na população: quando as soluções eram infactíveis, as soluções seriam escolhidas de acordo com o nível de dominância em relação à violação das restrições, dando uma ênfase maior às soluções que não satisfaziam o maior número de restrições. Porém, à medida que a população passava a conter apenas soluções factíveis, o algoritmo tenderia à estagnar.

Em (Deb et al., 2000), os autores modificam o princípio de dominância: antes de comparar duas soluções com relação à dominância, a factibilidade das soluções era verificada. Se uma solução era factível e a outra não, a solução factível dominava a infactível. Se as duas soluções eram infactíveis, a solução com menor violação de restrições dominava a outras. Se ambas as soluções eram factíveis, aplicava-se o princípio de dominância usual. Esse procedimento proposto não necessita de nenhum outro método de tratamento de restrições e pode ser utilizado em qualquer algoritmo evolutivo multiobjetivo.

Esta abordagem, utilizada com o NSGA-II, foi testada em um conjunto de problemas de teste ((Deb et al., 2001)), juntamente com as outras duas abordagens descritas acima ((Jiménez et al., 1999; Ray et al., 2000)) . Os resultados mostraram uma superioridade do primeiro método em relação aos outros dois.

Seguindo a mesma linha de operadores de busca local baseados em aproximações quadráticas de funções, em (Wanner et al., 2008), os autores propuseram um operador adicional capaz de encontrar estimativas Pareto-ótimas mais precisas. O novo operador usava aproximações quadráticas para as funções objetivo e de restrições, que eram construídas usando as avaliações de funções usuais do algoritmo evolutivo. A fase de busca local consistia em resolver um problema multiobjetivo quadrático, via uma escalarização usando uma formulação conhecida como *goal attainment* ((Ehrgott, 2000)). A solução desse problema era introduzida na população do algoritmo. Esta metodologia foi testada em problemas analíticos e em problemas onde a avaliação da função objetivo impunha um alto custo computacional e, em ambos os casos, os resultados indicaram que a metodologia era capaz de encontrar conjunto Pareto-ótimo de qualidade superior se comparados com um algoritmo multiobjetivo tradicional.

Apesar do grande número de trabalhos em otimização multiobjetivo, existem poucos trabalhos que lidam com tratamento de restrições nesse tipo de problema ((Coello Coello, 1999)). A maior parte dos trabalhos lidam com problemas irrestritos, usam penalidade estática ou tratam as restrições como funções objetivo adicionais.

## 7. Conclusões

Neste capítulo apresentamos diversas técnicas de tratamento de restrições em algoritmos evolutivos, mono e multiobjetivo. As técnicas apresentadas tratam desde os métodos de penalidade e suas variações à técnicas híbridas, combinando um algoritmo evolutivo com uma heurística ou um método numérico.

Todos os trabalhos nessa direção são, por assim dizer, inconclusivos. Mesmo que uma determinada metodologia funcione bem em uma certa classe de problemas, esta tenderá a ser inferior em outra. Dependendo do tipo de problema a ser abordado, novas técnicas precisaram ser desenvolvidas. Este é um campo vasto de pesquisa e há muito ainda a ser desenvolvido.

De um modo geral, as técnicas de tratamento de restrições devem levar em consideração os seguintes aspectos:

- serem genéricas, possibilitando pequenas modificações para lidar com o maior número de problemas;
- não necessitar de ajustes dos parâmetros, ou pelo menos minimizar esses ajustes, facilitando a utilização para o usuário;
- eficiência, não acarretando ao algoritmo um alto custo computacional na sua aplicação;
- incorporar o maior conhecimento possível sobre o domínio, melhorando assim o desempenho do algoritmo.

Uma técnica de tratamento de restrições ideal deveria incorporar todos esses apectos, entretanto sabe-se que, na prática, isso é impossível. Por exemplo, ao incorporar conhecimento sobre o domínio haverá naturalmente uma redução da generalidade da técnica ((Goldberg, 1989)). Apesar desses aspectos serem conflitantes, é necessário tê-los em mente ao propor uma técnica de tratamento de restrições. É ainda recomendável que as limitações de cada técnica proposta conhecida, facilitando a escolha frente a um determinado problema de otimização.