

MATLAB으로 배워 보는 유전 알고리즘 (2)

김재은

(대구가톨릭대학교 기계자동차공학부)

1. 머리말

필자는 지난 호에서 MATLAB에서 제공하는 유전 알고리즘의 기본 개념, Toolbox 구성 및 초기 집단 생성에 대해 설명한 바 있다. 이번 호에서는 선택, 교차(교배), 변이의 3가지 기본 연산에 대한 여러 선택 사항을 설명하고, 마지막으로 간단한 구조 최적화 문제에 대해 ANSYS 유한요소 해석 소프트웨어의 script(APDL) 파일과 m-file을 연동시켜 MATLAB 유전 알고리즘을 적용해 보았다.

2. 유전 알고리즘의 선택 사항 설명

2.1 적합도 조정

유전 알고리즘의 3가지 기본 연산 중 선택 연산은 다음 세대의 부모가 될 염색체들을 선택하기 위해 우선 각각의 염색체들에 대해 평가된 적합도 함수 값을 조정(scaling)한다. 우수한 개체(염색체)들과 열등한 개체들 사이의 적합도에 따른 선택 확률의 차이를 선택압(selection pressure)이라고 하는데 조정된 적합도의 분포가 크면 우수한 적합도를 가진 개체는 다시 선택되어질 확률이 커지는 반면, 그 분포가 작으면 각 개체별로 선택될 확률이 비슷해지므로 탐색이 느려질 수밖에 없다. 유전 알고리즘은 의미상 적합도가 큰 개체가 선택되어야 하므로 MATLAB 유전 알고

리즘에서는 사용자가 정의한 적합도 함수 f 를 최소화하기 위해, 가장 작은(큰) 적합도 함수값을 갖는 개체는 적합도 조정 결과에 의해 가장 큰(작은) 적합도 조정 결과값을 갖도록 되어 있다. 또한, 모든 조정된 적합도 값은 0 이상의 값을 갖도록 되어 있으며, 전체 합은 부모 개체의 수와 같도록 조정된다.

(1) 순위 기반 적합도 조정

MATLAB 유전 알고리즘의 기본 적합도 조정 방법으로서 각 개체의 순위에 기반하여 다음과 같이 사용자가 정의한 적합도 함수를 조정한다.

- 원래의 적합도 값이 작은 순서대로 n 번째 순위를 갖는 i 세대의 개체 k 의 조정된 적합도 \tilde{f}_k^i 는 $1/\sqrt{n}$ 에 비례한다.
- 각 개체들의 조정된 적합도 \tilde{f}_k^i 는 전체 합이 다음 세대 생성을 위한 부모 개체들의 수와 같도록 다음과 같이 주어진다(N 은 집단(population)의 크기).

$$\tilde{f}_k^i = \text{부모 개체의 수} \times \frac{1}{\sqrt{n}} / \sum_{n=1}^N \frac{1}{\sqrt{n}} \quad (1)$$

(2) 비례 적합도 조정

MATLAB 유전 알고리즘에서 제공하는 가장 간단한 적합도 조정 방법으로서 각각의 개체가 갖는 원래의 적합도 함수 값에 비례하여 조정을

한다. 이 방법은 원래의 적합도 값들이 적절한 범위에 있지 않으면 좋은 효과를 거둘 수 없다. 이 선택 사항에서는 가장 작은 원래의 적합도 함수 값이 가장 큰 조정값을 갖도록 다음과 같이 재계산된다.

$$\tilde{f}_k^i = \frac{2}{N} \sum_{k=1}^N f_k^i - f_k^i \quad (2)$$

만약, 위의 식에서 음의 값이 나오는 경우 가장 작은 값을 구하여 그 값이 0이 되도록 $\tilde{f}_k^i = f_k^i - \min_i(\tilde{f}_k^i)$ 을 이용하여 재조정한다. 최종 조정된 적합도 값은 다음과 같다.

$$\tilde{f}_k^i = N \times \tilde{f}_k^i / \sum_{k=1}^N \tilde{f}_k^i \quad (3)$$

(3) 최상위 개체 적합도 조정

MATLAB 유전 알고리즘에서의 적합도 조정을 'Top' 으로 선택하면 'Quantity' 라는 선택 사항을 제공한다(기본값은 0.4). 여기에서 1과 집단의 크기 사이의 값 또는 0과 1사이의 값을 입력하면, 원래의 적합도 값이 작은 순서대로 Quantity (0과 1사이의 값이 주어진 경우 Quantity=Quantity×집단의 크기)에서 입력한 숫자만큼의 개체에 조정된 적합도 N/Quantity가 주어지고 나머지 개체에 0이 부여된다.

(4) 선형 이동 적합도 조정

목적 함수의 값이 가장 작은 개체가 가장 큰 적합도 함수 값을 갖도록 조정하므로 우선 중간 단계로서 각 개체에 대한 원래의 적합도 함수 값에 모두 -1을 곱하여 $-f_k^i$ 를 만들고, $\tilde{f}_k^i = a \times (-f_k^i) + b$ 가 되도록 적합도 함수를 조정한다. a 와 b 의 값은 추가로 보여지는 사용자 입력 값 'Max survival rate' (기본값 2)를 이용하여 다음을 만족하도록 구해진다.

- $a \times \max(-f_k^i) + b = \text{Max survival rate} \times V_{avg}$
(V_{avg} = 부모 개체들의 수/k의 개수)
- $a \times \text{mean}(-f_k^i) + b = V_{avg}$

이 조정법은 초기 세대에서 초우량 개체의 영향을 줄이고, 후 세대에서는 각 개체간 선택 확률 차이를 넓게 해주는 역할을 한다.

2.2 선택 연산

자연 환경하에서의 생물은 생존 경쟁을 통해 남거나 도태되는 자연 선택(natural selection)의 과정을 통해 생물학적인 진화를 경험한다. 이를 모방하기 위한 유전 알고리즘 기본 연산을 '선택' 이라고 하며, 다음 자식 세대를 재생산하기 위한 교차(교배) 및 변이 연산자의 대상이 되는 개체를 생성한다. 여기서는 'Options-Selection' 에서 보여주는 선택 연산자의 순서를 따르지 않고 기능이 단순한 것부터 설명하기로 한다.

(1) 균등 선택

사용자가 정의한 선택 함(custom)의 debugging 을 위해 제공되는 것으로서, 실제 문제에는 적용하지 않도록 한다.

(2) 룰렛 휠 선택

이 방법은 선택 연산자의 선택 사항에 나열된 것 중 4번째에 위치하고 있으나 유전자 알고리즘에 관한 책에서는 일반적으로 가장 먼저 설명되는 선택 연산자이고 또한 다른 선택 연산자를 설명함에 있어서 비교 대상이 되므로 여기서는 먼저 설명하기로 한다.

앞서 기술한 바와 같이 MATLAB 유전 알고리즘에서는 조정된 적합도 함수 값 전체의 합이 부모 개체의 수와 같도록 조정되므로, 각각의 개체

기초강좌

는 해당 개체의 조정된 적합도 값/부모 개체의 수와 같은 면적을 룰렛 휠 상에 할당 받게 된다. 최종 누적 확률을 계산하고 (0, 1) 사이의 수를 무작위로 발생시켜 개체를 선택하게 된다. 예를 들면, i 세대의 개체를 $(s_1(i), s_2(i), s_3(i), s_4(i), s_5(i), s_6(i))$ 라고 하고, 각각의 해당 룰렛 휠 면적이 0.21, 0.39, 0.13, 0.14, 0.08, 0.05라고 하자. 이에 대한 누적 선택 확률은 $P_1 = 0.21, P_2 = 0.60, P_3 = 0.73, P_4 = 0.87, P_5 = 0.95, P_6 = 1.00$ 이 된다. 따라서, (0,1)에서 무작위로 추출된 수 r 이 $P_{i-1} \leq r < P_i$ ($i = 1 \sim 6, P_0 = 0$)을 만족하면, $s_k(i)$ 를 선택하게 된다. 즉, 발생된 난수가 [0.4387, 0.3816, 0.7655, 0.7952, 0.1869, 0.4898] 이라면, 자식 세대 생성을 위한 교차 및 변이 연산 대상의 개체들은 $(s_2(i), s_2(i), s_4(i), s_4(i), s_1(i), s_2(i))$ 가 된다.

(3) 확률 균등 선택

MATLAB 유전 알고리즘의 기본 선택 연산자로서 각 개체의 누적 선택 확률을 이용하여 선택하는 점은 룰렛 휠 방법과 유사하다. 그러나, (0,1) 구간에서 발생된 난수를 이용하여 선택하는 대신, (0,1/부모개체의 수) 구간에서 균일 분포 난수 1개를 발생시키고, 이 난수로부터 부모 개체의 수만큼 구간의 간격 크기를 누적 합산한 숫자들을 이용한다. 즉, 위의 룰렛 휠 예제에서 (0,1/6) 구간에서 생성된 균일 분포 난수가 0.0743 이라면, [0.0743, 0.2410, 0.4076, 0.5743, 0.7410, 0.9076]를 얻게 되고, 이를 각 개체의 누적 선택 확률에 적용하면 $(s_1(i), s_2(i), s_2(i), s_2(i), s_4(i), s_5(i))$ 가 선택된다.

(4) 잔여 확률 표본 선택

각 개체의 조정된 적합도의 정수 값만큼 해당 개체를 우선 선택하고 나머지는 소수 값들을 이

용하여 확률 균등 방법에 의해 선택한다.

(5) 토너먼트 선택

추가 입력 사항으로서 'Tournament size'에서 2와 N (N : 집단의 크기) 사이의 값을 입력 받아 이 크기로 임의의 개체를 선택하고, 그 중에서 가장 적합한 부모 1개체를 선택한다. 이렇게 하여, N 개의 개체가 선택될 때까지 반복된다. 토너먼트 크기를 크게 할수록 선택압이 증가한다. 기본값은 4로 설정되어 있다.

2.3 교배/교차 연산

2개의 부모 세대를 교배하여 1개의 자식 세대를 생성하기 위한 연산으로서 정수 문제(integer problems)에는 사용하지 않는다. 다음에 설명할 일점 및 이점 교배 등 전통적인 교차 연산자는 2개의 부모 개체로부터 2개의 자식이 생성되지만 MATLAB 유전 알고리즘에서는 모든 교차 연산자에 대해 2개의 부모 개체로부터 1개의 자식이 생성되고 부모 개체의 수가 될 때까지 연산이 반복된다.

(1) 산란 교차

MATLAB 유전 알고리즘에서 제공하는 기본 교차 연산자로서 임의의 이진수 벡터를 생성한 후 하나의 부모로부터는 '1' 위치의 유전자를 선택하고, 다른 하나의 부모로부터는 '0' 위치의 유전자를 선택한다. 예를 들어, 부모 염색체가 각각 $s_1(i) = [a, b, c, d, e, f, g, h]$ 및 $s_2(i) = [1, 2, 3, 4, 5, 6, 7, 8]$ 이고 임의로 생성된 이진수 벡터가 [1, 1, 0, 0, 1, 0, 0, 0]이라면, 자식은 $\tilde{s}(i+1) = [a, b, 3, 4, e, 6, 7, 8]$ 로서 생성된다.

(2) 일점 교차

우선 1과 설계 변수 개수 사이에서 정수인 난수

를 생성한다. 교차 연산 대상 부모 개체 중 1개의 개체에서는 그 난수보다 작거나 같은 위치에 있는 유전자 군을 선택하고, 다른 1개의 부모 개체에서는 큰 위치에 있는 유전자 군을 선택하여 서로 결합한다. 예를 들어, 부모 염색체가 각각 산란 교차에서의 예제와 같고 1과 8사이에서 임의로 생성된 난수가 3이라면 (이를 교배점 또는 교차점이라고 한다.) 자식은 $\tilde{s}(i+1)=[a, b, c, 4, 5, 6, 7, 8]$ 이 된다.

(3) 이점 교차

임의로 생성된 2개의 교배점 m 과 n ($m > n$)이 이용된다. 위의 일점 예제에서 3과 6의 교배점을 적용하면, 자식은 $\tilde{s}(i+1)=[a, b, c, 4, 5, 6, g, h]$ 가 된다.

(4) 중간 교차

추가 입력 사항으로 제공되는 'Ratio' (기본값은 1)에 상수 또는 설계 변수와 같은 크기의 행벡터를 지정함으로써 자식 개체는 다음과 같이 교차 대상 부모 개체의 가중치 평균을 취한다.

$$\tilde{s}(i+1) = s_k(i) + \text{rand} \times \text{Ratio} \times (s_{k+1}(i) - s_k(i)) \quad (4)$$

(rand: 0과 1사이의 균일 분포 난수)

(5) 학습 교차

자식 세대는 2개의 부모 세대를 포함하는 선상에 존재하나, 2개의 부모 세대 중 더 좋은 적합도를 갖는 부모 세대에 근접하도록 생성된다. 자식 개체가 더 좋은 적합도를 갖는 부모 개체에 얼마나 떨어져 있어야 하는가는 Ratio를 통해 조절한다(기본값은 1.2). 즉, $s_{k+1}(i)$ 가 $s_k(i)$ 보다 더 좋은 적합도를 갖는다면 자식 개체는 다음에 의해 생성된다.

$$\tilde{s}(i+1) = s_k(i) + \text{Ratio} \times (s_{k+1}(i) - s_k(i)) \quad (5)$$

(6) 산술 교차

(0,1) 구간에서 생성된 난수 α 를 이용하여 2개 부모 세대의 가중치 산술평균을 이용한다. 즉, $s_k(i)$, $s_{k+1}(i)$ 에 대해 $\tilde{s}(i+1) = \alpha s_k(i) + (1 - \alpha) s_{k+1}(i)$ 가 된다.

2.4 변이 연산

변이 연산자는 변이 연산 대상 부모 개체들에 대해 미소한 무작위 변화를 가하여 유전적 다양성을 제공하고 유전 알고리즘으로 하여금 좀 더 넓은 설계 공간을 탐색하게 한다. 정수 문제에는 사용하지 않는다. 기본 선택은 'constraint dependent'이며, 이는 경계 조건이 없으면 정규 변이를, 그 외에는 설계 영역 적용 변이를 기본으로 적용함을 의미한다.

(1) 정규 변이

MATLAB 유전 알고리즘에서의 기본 변이 연산자이다. 이는 변이 연산 대상이 되는 부모 벡터의 각 유전자에 표준 정규 분포로부터 임의로 추출된 난수를 더하게 된다. 변이 양은 분포의 표준 편차에 비례하며 이는 각 세대를 거치면서 감소한다. 분포의 표준편차는 'Population' 선택 란의 'Initial range' 및 추가 입력 사항 'Scale', 'Shrink'에 의해 결정된다.

- 'Scale'은 초기 세대에서의 표준 편차를 결정한다. 즉, 초기 변이 연산 대상 부모 개체의 각 유전자에 더해지는 난수는 표준편차의 값이 $\text{Scale} \times (\text{'Initial range'에서 의해 설정된 상한값과 하한값의 차이})$ 를 갖는 표준 정규 분포에서 추출된다.

- 'Shrink'는 세대가 진화하면서 표준편차 및 변이가 감소되는 방법을 제어한다. 표준 편차는

기초강좌

최종값이 1-Shrink가 되도록 선형으로 감소한다. 따라서, 'Shrink'의 값을 1로 설정하면 각 세대에서 난수 추출 대상 분포의 표준편차는 마지막 세대에서 0이 될 때까지 선형으로 감소한다. 'Scale'과 'Shrink'의 기본값은 1이며, 음의 값을 가지면 세대가 진화할수록 표준편차는 증가한다.

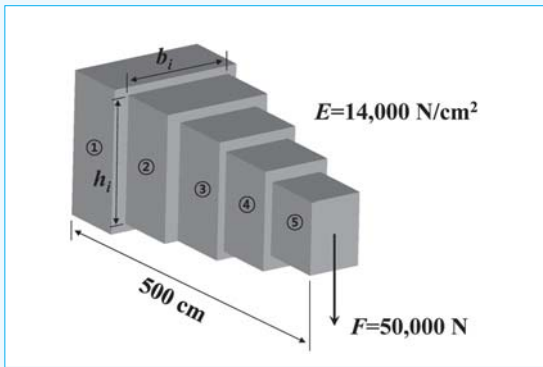


그림 1 유전 알고리즘 최적화 예제(외팔보)

(2) 균등 변이
변이 연산 대상 부모 세대에서 각 유전자는 0과 1사이의 균등 분포에서 취해진 확률이 할당된다. 추가 입력 사항인 'Rate'에서 설정된 값(기본값은 0.01) 보다 작은 유전자는 'Initial range'에 해당하는 균등 분포에서 추출된 값으로 대체된다.

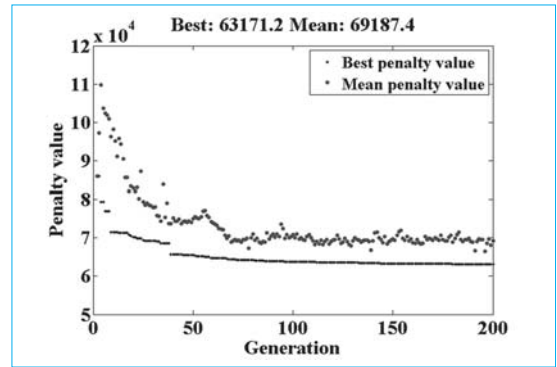


그림 3 각 세대별 적합도 함수의 평균, 최적 값

목적함수 (적합도) 계산

```
function object=Cal_Obj(dv)
dv=dv';
fid = fopen('design_var.dat','wt');
fprintf(fid,'%10.5f %n',dv);
fclose(fid);
유한요소해석 script file 실행문
load object.dat -ascii
```

제한 조건 계산

```
function [cnst,cnsteq]=Cal_Cnst(dv)
dv=dv';
fid = fopen('design_var.dat','wt');
fprintf(fid,'%10.5f %n',dv);
fclose(fid);
유한요소해석 script file 실행문
load cnst.dat -ascii
cnsteq=[];
```

설계 변수의 개수

정수 설계 변수 지정

설계 변수 하한, 상한 지정

그림 2 GUI 입력 변수 지정 및 m-file과의 연동

(3) 설계 영역 적응 변이
경계 조건이 있는 경우의 기본 변이 함수로서 각 설계 변수의 상/하한 경계와 선형 제한 조건을 만족하도록 방향과 변이 증분을 조절한다.

3. 유전 알고리즘의 적용

지금까지 설명한 MATLAB 유전 알고리즘을 그림 1에 나타낸 외팔보 구조 최적화 문제에 적용해 보고자 한다. 이 예제는 MATLAB 'Global optimization toolbox' 사용자 매뉴얼에 나와있지만(R2012b부터 추가됨), 여기서는 동일한 문제를 ANSYS script 파일과 MATLAB 유전 알고리즘을 연동하여 풀고자 한다. 목적 함수 및 제한 조건 등은 다음과 같다.

$$\text{Minimize } 100 \times \sum_i^5 b_i h_i \quad (6)$$

subject to $\delta_{ip} \leq 2.75 \text{ cm}$

$$\sigma_i \leq 14,000 \text{ N/cm}^2$$

$$1 \leq b_1 \leq 5; 30 \leq h_1 \leq 65$$

$$2.4 \leq b_2, b_3 \leq 3.1; 45.0 \leq h_2, h_3 \leq 60.0$$

$$1.0 \leq b_4, b_5 \leq 5.0; 30.0 \leq h_4, h_5 \leq 65.0$$

위 식에서 δ_{ip} 은 끝단의 처짐을, σ_i 는 i 번째 구간(각 구간의 길이는 100 cm)의 최대 굽힘 응력을 각각 나타낸다. 또한, 이 문제의 특징은 첫 번째

구간 계단 보의 단면 높이(h_1) 및 폭(b_1)이 정수값을 갖는 혼합 정수 설계 변수 최적화 문제이다. 그림 2는 GUI 창의 입력 변수 값 및 함수 m-file 구성을 나타낸다[사용된 ANSYS script 파일은 필자(jekim@cu.ac.kr)에게 연락하면 개인적으로 제공받을 수 있다.]. 예제를 풀기 위해 여기서는 기본 설정 이외의 사항으로서 초기 집단의 크기, 엘리트 개체수를 각각 150 및 10으로 설정하였다. 또한, 전체 진화 과정을 200세대로 한정하였다. 그림 3은 'Plot functions' 항목을 'Best fitness'로 설정하였을 때 각 세대에서의 최적 및 평균 적합도 값을 보여주고 있다. 최적화 결과로서 각 설계 변수의 값이 ($b_1, h_1, b_2, h_2, b_3, h_3, b_4, h_4, b_5, h_5$) = (3, 60, 2.85, 57.06, 2.60, 51.01, 2.19, 43.81, 1.78, 33.83) (cm)일 때 전체 계단보의 체적은 63171.2 cm³가 된다.

4. 맺음말

지금까지 총 2회에 걸쳐 MATLAB 유전 알고리즘에 대해 살펴보는 시간을 가졌다. 대부분의 기술은 주로 MATLAB 사용자 매뉴얼 및 직접 m 파일을 분석한 결과를 기초로 하였음을 밝혀둔다. 지면의 제약상 Toolbox GUI의 다른 내용은 일부 생략하였지만, 대부분은 설정된 기본값을 사용하여도 무방하다. 더욱 자세한 내용은 MATLAB 사용자 매뉴얼을 참고하기 바란다. **KSNVE**