Obliczenia z wykorzystaniem sztucznej inteligencji

wykład II
Strategie ewolucyjne

Joanna Kołodziejczyk

Plan wykładu

Jak powstały Strategie Ewolucyjne

Autorzy:

Ingo Rechenberg i Hans-Paul Schwefel Technische Universitat Berlin.

Cel:

Opracować zestaw reguł potrzebnych do automatycznego projektowania poprzez krokowe eksperymentowanie z dopasowywaniem zmiennych, prowadzące do tego, że modelowany obiekt/system osiągnie swój stan optymalny niezależnie od zaszumienia środowiska.

Reguły:

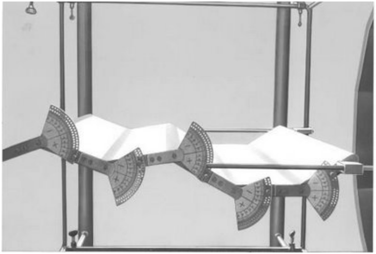
Reguła 1: Zmień wszystkie zmienne w tym samym czasie, najlepiej losowo i nieznacznie.

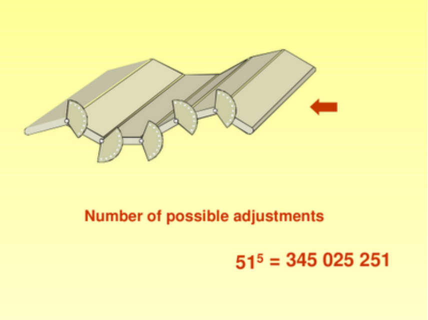
Reguła 2: Jeżeli nowy zestaw zmiennych nie zmniejsza jakości urządzenia/systemu, zachowaj go, w przeciwnym wypadku zachowaj stary zestaw.

Eksperyment dla (1+1)-ES

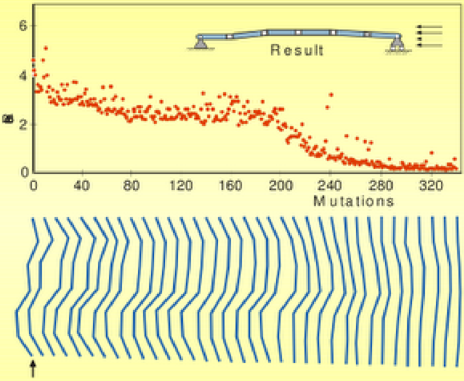
Cel eksperymentu:

1. Udowodnić zbieżność metody.
2. Metoda nie wymaga zbyt wielu powtórzeń, by osiągnąć optymalne (lub zadowalające rozwiązanie)



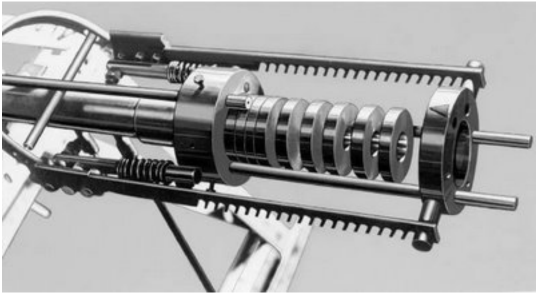


Wyniki eksperymentu



The *experimentum crucis* - Drag minimization of the kink plate

Eksperyment z dyszą dwufazową



http://ls11-www.cs.uni-dortmund.de/people/schwefel/EADemos/

Przeznaczenie ESs

Cel:

Optymalizacja (pewnej) zadanej funkcji celu lub funkcji jakości F (zazwyczaj) wielu zmiennych decyzyjnych lub parametrów sterujących $\mathbf{y} := (y_1, y_2, \dots)$ nazywanych *parametrami celu*.

$$F(\mathbf{y}) \to opt., \mathbf{y} \in \mathbb{Y}$$

$\mathbb{Y}$ może być skończonym wektorem dowolnych danych, ale niekoniecznie stałej długości, np.:

- liczby rzeczywiste w N-wymiarowej przestrzeni $\mathbb{R}^N$
- liczby całkowite $\mathbb{Z}^N$
- liczby binarne $\mathbb{B}^N$
- permutacje $\mathbb{P}^N$
- inne, dowolne przestrzenie mieszane

Populacja, osobnik

Populacja

Niech $\mathfrak{P}$ oznacza populację osobników $\mathfrak{a}$.

Osobnik

$$\mathfrak{a}_k := (\mathbf{y}_k, \mathbf{s}_k, F(\mathbf{y}_k))$$

gdzie:

- $\mathbf{y}_k$ — wektor parametrów
- $F(\mathbf{y}_k)$ — funkcja celu zwana też przystosowaniem
- $\mathbf{s}_k$ — endogenne parametry strategii — odróżnia ESs od innych algorytmów — używane do kontrolowania własności statystycznych operatorów rekombinacji, najczęściej mutacji. Poddawane są procesowi ewolucji wraz z innymi parametrami.

Plan wykładu

1. Koncepcja ES

2. Algorytm
   - Kroki algorytmu
   - Parametry
   - Selekcja
   - Mutacja
   - Rekombinacja
   - Ewolucja parametrów endogennych

3. Literatura

Algorytm $(\mu/\rho \overset{+}{,} \lambda)$-ES

1. Inicjalizacja populacji początkowej $\mathfrak{P}$ składającej się z μ
2. Utwórz λ potomków, tworząc populację potomków, gdzie każdy potomek powstaje poprzez wykonanie:
   2.1 Wybierz losowo ρ rodziców z populacji $\mathfrak{P}$
   2.2 Utwórz potomka *r* w wyniku wykonania operacji rekombinacji na parametrach endogennych, a potem na wektorze parametrów.
   2.3 Dokonaj mutacji potomka *r*, wykonując mutację parametrów endogennych, a potem wektora parametrów.
3. Wybierz nową populację poprzez jedną z selekcji
   3.1 typu comma , wybór z populacji potomków
   3.2 typu plus $+$ wybór z populacji potomków i rodziców
4. Idź do kroku 2, dopóki nie jest spełnione kryterium zakończenia działania algorytmu

Parametry egzogenne

ρ

Liczba rodziców zaangażowanych w proces rekombinacji tworzącej **jednego** potomka.
$\rho = 1$ — klonowanie bez rekombinacji zapisywane jako (μ, λ)-ES lub $(\mu + \lambda)$-ES.
$\rho > 1$ — strategia z rozmnażaniem

μ

Liczność populacji rodzicielskiej (w tym wielkość populacji początkowej).

λ

Liczba potomków, tj. tymczasowej populacji potomnej powstałej przez działanie operatorów rekombinacji i mutacji na populacji rodziców o liczności μ.

UWAGA: Parametry egzogenne są niezmienne, zadawane jako parametr.

Selekcja w ES

Selekcja

W przeciwieństwie do losowych operatorów rekombinacji jest ukierunkowana na osiągnięcie celu optymalizacji.

breeding/truncation selection

W ESs selekcja jest *deterministyczna*. Gwarantuje, iż tylko μ najlepszych osobników z puli selekcyjnej pewnej generacji g o liczności γ zostanie przetransferowana do populacji potomnej.

Typy selekcji

1. $(\mu + \lambda)$ — zarówno rodzice jak i potomkowie są włączeni do procesu selekcji, czyli $\gamma = \mu + \lambda$. Nie ma żadnych ograniczeń nałożonych na μ i λ.
   Wariant $(\mu + 1)$ nazywany jest **steady-state ES** i jest implementowany w systemach wieloprocesorowych.
   Selekcja jest **elitarna**, tj. przechowuje zawsze dotychczas znalezionego osobnika.
   Zastosowania: w przestrzeniach dyskretnych skończonych.
2. (μ, λ) — tylko potomkowie biorą udział w selekcji, czyli $\gamma = \lambda$. Rodzice są zapominani, czyli **brak elitaryzmu**.
   Dla zachowania zbieżności algorytmu konieczny jest warunek, by $\mu < \lambda$.
   W przypadku, gdy $\mu = \lambda$ algorytm przestaje być kierunkowy i losowo przemierza przestrzeń poszukiwań.
   Zastosowania: nieskończone przestrzenie poszukiwań, w szczególności w $\mathbb{R}^N$

Mutacja — główny operator ES

Brak ustalonej metodologi doboru operatora mutacji, który jest problemo-zależny. Kilka zasad zostało jednak określonych:

1. reachability — z podanego stanu rodzicielskiego $(\mathbf{y}_p, \mathbf{s}_p)$ w skończonej liczbie generacji osiągnięty zostanie stan potomny $(\tilde{\mathbf{y}}, \tilde{\mathbf{s}})$ (warunek konieczny dowodzenia globalnej zbieżności).
2. unbiasedness — Jeżeli selekcja jest ukierunkowana, to nie ma potrzeby nakładania warunku na operatory rekombinacji i mutacji. Prowadzi to do zasady maksymalnej entropii co w konsekwencji dla przestrzeni poszukiwań $\mathbb{R}^N$ daje rozkład normalny, a dla $\mathbb{Z}^N$ rozkład geometryczny.
3. scalability — średnia długość kroku mutacji powinna dostrajać się do własności krajobrazu przystosowawczego (fitness landscape), które zależą od funkcji celu i operatorów rekombinacji i mutacji.
   Koncepcja współzależności (Rechenberg '95): małe zmiany na poziomie genetycznym powinny powodować niewielkie zmiany w wartości przystosowania.

Mutacja parametrów **y**— przestrzeń $\mathbb{R}^N$

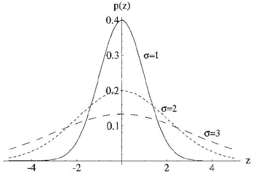
Zakłada się, że siła mutacji określona jest odchyleniem standardowym σ i stanowi jedyny parametry ESs w wektorze **s**.

Isotropic mutation

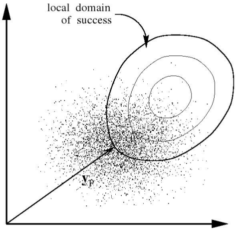
$$\tilde{\mathbf{y}} = \mathbf{y} + \mathbf{z},$$

gdzie: $\mathbf{z} = \sigma(\mathcal{N}_1(0,1), \ldots, \mathcal{N}_N(0,1))$

Mutacja jest równomiernie rozłożona wzdłuż wszystkich N osi
Zalety: Jest tylko jeden parametr sterujący, który należy kontrolować.



a) 1-D $p(z)$ density curve



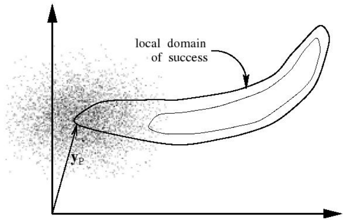
b) 2-D mutation samples

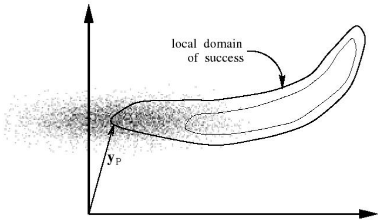
Mutacja parametrów **y**— przestrzeń $\mathbb{R}^N$ cd.

Gaussian mutation

$$\mathbf{z} = (\sigma_1\mathcal{N}_1(0,1), \ldots, \sigma_N\mathcal{N}_N(0,1))$$

Wektor adaptujących się endogennych parametrów będzie miał postać: $\mathbf{s} = (\sigma_1, \ldots, \sigma_N)$





Mutacja parametrów $\mathbf{y}$ — przestrzeń $\mathbb{R}^N$ cd.

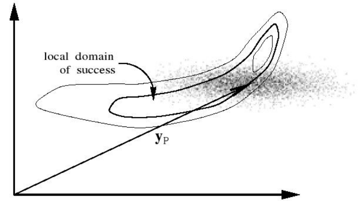
Correlated Gaussian mutation

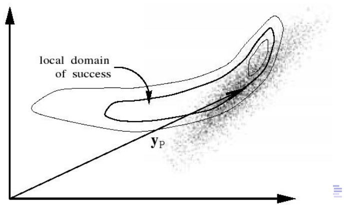
Uwzględnia również obrót w przestrzeni stanów.

$$\mathbf{z} = \mathbf{M}(\sigma_1 \mathcal{N}_1(0,1), \ldots, \sigma_N \mathcal{N}_N(0,1))'$$

gdzie: **M** jest ortogonalną macierzą rotacji przedstawiająca korelacje pomiędzy elementami **z**.

Wektor adaptujących się endogennych parametrów będzie zawierał: $N(N+1)/2$ elementów





Mutacja parametrów $\mathbf{y}$ — przestrzeń $\mathbb{B}^N$

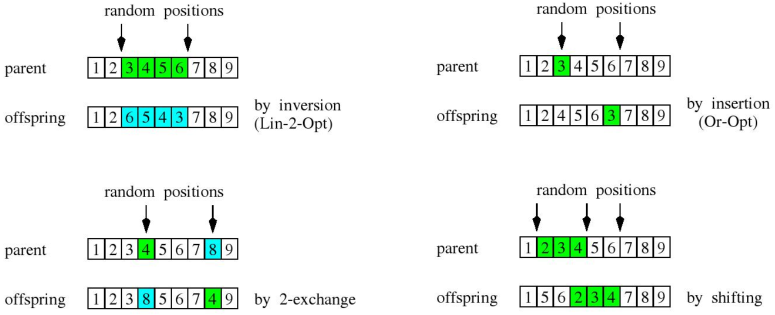
Binarry mutation

Mutacja polega na losowej zmianie wartości wektora $\mathbf{y}$.
Prawdopodobieństwo mutacji (p_m) postrzegane jest jako siła mutacji.

Zazwyczaj każdy gen jest mutowany niezależnie. Badania empiryczne jak i kilka teoretycznych dowodów sugerują wybranie $p_m = 1/N$ (Back '96). p_m może być zmieniane, np. (Jansen & Wegener) można zmieniać $p_m = f(g)$ cyklicznie z $1/N \leq p_m \leq 1/2$, co usprawni działanie $(1+1)$-ES dla pseudoboolowskich funkcji celu.

Można wprowadzić bardziej złożone operacje binarne oraz rozkłady zmiennych boolowskich (*factorized distribution algorithm* czy *Bayesian optimization algorithm*).

Mutacja parametrów **y** — przestrzeń kombinatoryczna

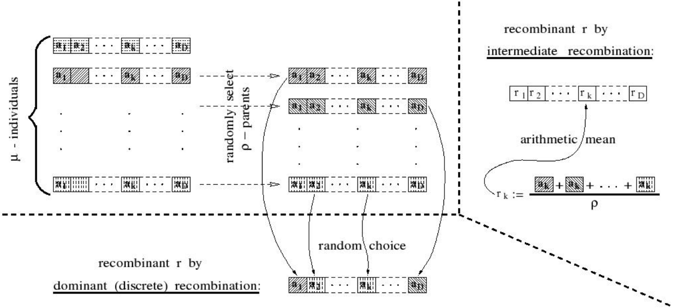


Neighborhood

Liczba stanów możliwych do osiągnięcia ze stanu rodzicielskiego w jednym kroku jest siłą mutacji w przestrzeni permutacyjnej. Jako że neighborhood zazwyczaj jest stałe, to zmiennym parametrem jest liczba mutacji przeprowadzana w pojedynczym kroku. $1 < s < const \cdot N$.

Rekombinacja parametrów **y** przestrzeń dyskretna i ciągła

- Rekombinacja wykorzystuje informację z ρ rodziców.
- $\rho > 2$ multirekombinacja.
- Rekombinacja produkuje 1 potomka.



Dyskusja rekombinacji

Building block hipothesis (Goldberg '89)

Rekombinacja pomiędzy dobrymi rodzicami prowadzi do powstania dobrego potomka (intuicja). '01 Jansen & Wegener stworzyli funkcję potwierdzającą hipotezę dla krzyżowania jednopunktowego.

GR - genetic repair hipthesis (Beyer '97)

Wspólne cechy (podobieństwa) przechodzą do potomka, a nie jak w BBH cechy oczekiwane i różnorodne.
Należy założyć, że odpowiednie elementy genomu rodziców, które są do siebie podobne niosą większe prawdopodobieństwo bycia przydatnymi w odniesieniu do przystosowania (pochodzą z wybranych, najlepiej przystosowanych rodziców). Zatem operator powinien zgodnie z zasadą maksymalnej entropii zachować, cechy powtarzalne, a inne jako mniej istotne zmieniać losowo.

Dyskusja rekombinacji cd.

W $\mathbb{R}^N$ multirekombinacja z mutacją jednoparametryczną $(\mu/\mu, \lambda)$-ES

Powstanie potomek

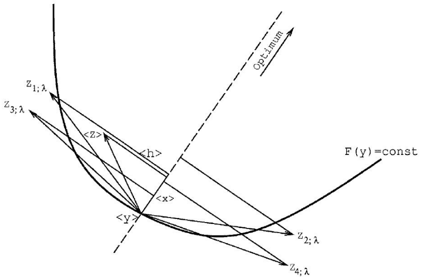
$$\langle \mathbf{y} \rangle^{(g+1)} = \langle \mathbf{y} \rangle + \frac{1}{\mu} \sum_{m=1}^{\mu} \mathbf{z}_m,$$

„The new parental centroid is obtained from the old one by adding the vector average $\langle \mathbf{z} \rangle$ of the μ best mutations."

Ilustracja efektu Genetic Repair w $\mathbb{R}^N$

Dekomponujemy $\mathbf{z}$ na składowe: $\mathbf{z} = x\mathbf{e}_{opt} + \mathbf{h}$, gdzie $\mathbf{h}$ — składowa niekorzystna, a $x\mathbf{e}_{opt}$ — składowa przesuwająca w kierunku optimum.

$$\langle y\rangle^{(g+1)} = \langle y\rangle + \langle z\rangle = \langle y\rangle + \mathbf{e}_{opt} + \frac{1}{\mu}\sum_{m=1}^{\mu} x_{m;\lambda} + \frac{1}{\mu}\sum_{m=1}^{\mu} \mathbf{h}_{m;\lambda} = \langle y\rangle + \langle x\rangle\mathbf{e}_{opt} + \langle h\rangle$$



Genetic Repair w $\mathbb{R}^N$ cd.

- Hans Beyer w'95 wykazał, zakładając statystyczną niezależność elementów wektora **h**, iż oczekiwana długość $\langle \mathbf{h} \rangle$ będzie mniejsza o faktor $1/\sqrt{\mu}$ niż oczekiwana długość pojedynczego wektora **h**.
- „ This is the performance increasing effect of genetic repair, it reduces the fitness decreasing effect of the harmful parts of the mutants, thus, also allowing for a larger mutation strength and therefore a larger improvement step. To summarize, by conserving the beneficial parts of the mutations and dampening their harmful parts through genetic repair, intermediate recombination provides an efficient mechanism for similarity extraction in real-valued search spaces."

Ze względu na trudność zdefiniowania podobieństwa między wektorami rodzicielskimi w przestrzeni kombinatorycznej jak dotąd nie potwierdzono działania efektu GR w tej przestrzeni.

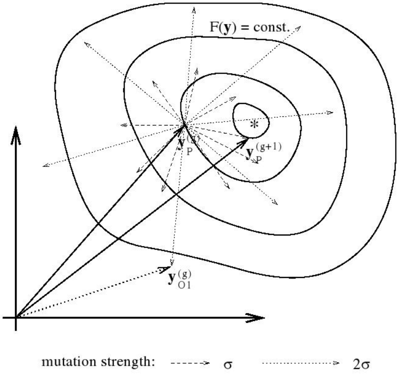
Adaptacja parametrów endogennych **s**

W ESs adaptacji ulegają parametry mutacji. Ich kontrolowanie odbywa się poprzez analizę statystyczną wprowadzanych przez nie zmian.
Wykorzystywane techniki ewolucji tych parametrów

- Reguła $1/5th$ dla strategii $(1+1)$-ES
- Techniki adaptacji oparte na informacji z lokalnej populacji
- Techniki zaawansowane z wykorzystaniem informacji z nielokalnej przestrzeni poszukiwań

Mutacja isotropic w $(1+1)$-ES

Krajobraz przystosowawczy w $\mathbb{R}^2$ dany przez linie stałej wartości $F(\mathbf{y})$.
$*$ - optimum



a) Mutations in an $\mathbb{R}^2$ search space.

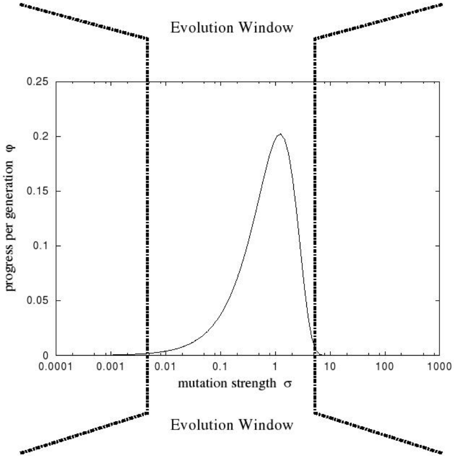
Jak σ wpływa na ESs

Oczekiwana długość mutacji jest proporcjonalna do σ

- małe σ — średnio co druga mutacja będzie trafiać w kierunek optimum:
  $\sigma \to 0: \quad P_s \to 1/2 \quad \textit{ale} \quad \varphi \to 0$
  gdzie: P_s — prawdopodobieństwo sukcesu (potomek wymienia rodzica), φ — współczynnik postępu (oczekiwane przesunięcie w kierunku optimum).
  „That is, even though we have a high degree of evolvability for sufficiently small mutation strengths, the performance of the ES will be rather poor. ”
- duże σ — szansa na trafienie w okolicę optimum maleje:
  $\sigma \to \infty: \quad P_s \to 0 \quad \textit{ale} \quad \varphi \to 0$

Evolution window

Badanie eksperymentalne wyznaczyły robocze zakresy wartości σ.

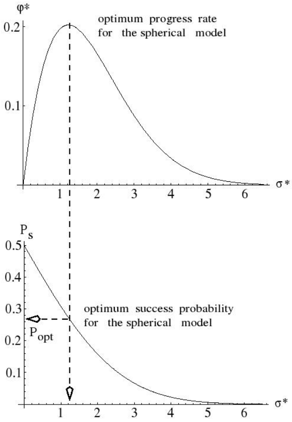


b) Progress rate and "evolution window."

Reguła $1/5th$ H. Rechenberg

- Stosowana w $(1+1)$-ES.
- Testowane na zadaniu optymalizacji funkcji sferycznej i liniowej (corridor), dało wyniki, że prawdopodobieństwo osiągnięcia sukcesu było bliskie 0.2
- „In order to obtain nearly optimal (local) performance of the $(1+1)$-ES in $\mathbb{R}^N$, tune the mutation strength in such a way that the (measured) success rate is about $1/5$."
- Jeżeli $P_s < 1/5$ $\sigma\downarrow$ lub $P_s > 1/5$, $\sigma\uparrow$.
- Współczynnik a zależy od F — funkcji celu, N — wymiarowości problemu i liczby G — liczby dobrych mutacji. Jeżeli $N \geq 30$ to $G = N$. Schwefel sugerował $0.85 \leq a < 1$.

Reguła 1/5*th* H. Rehcenberg



1. perform the $(1+1)$-ES for a number G of generations:
   - keep σ constant during this period,
   - count the number G_s of successful mutations during this period
2. determine an estimate of P_s by
$$P_s := G_s / G \tag{19}$$
3. change σ according to
$$\sigma := \begin{cases} \sigma / a, & \text{if} \quad P_s > 1/5 \\ \sigma \cdot a, & \text{if} \quad P_s < 1/5 \\ \sigma, & \text{if} \quad P_s = 1/5 \end{cases} \tag{20}$$
4. goto 1.

Stosowalność reguły $1/5th$

„(...) the 1/5th-rule has its specific application domain:

- it is restricted to the application of one strategy parameter,
- the fitness landscape must obey certain properties in order to ensure that for sufficiently small mutation strengths $P_s > 1/5$ (otherwise one may observe premature stagnation (Schwefel, 1995)),
- it is usually used in $(1 + 1)$-ES only (other strategies have different optimal P_s)."

Inne sposoby adaptacji endogennej

Główna idea adaptacji endogennej.

- Parametry strategii **s** dobierane są indywidualnie do parametrów obiektu **y**.
- **s** mogą przechodzić rekombinację i zawsze są poddane mutacji.
- Zgodnie z selekcją deterministyczną, przeżyją tylko takie parametry, które dały najkorzystniejsze zróżnicowanie dla **y**.

Z powyższego wynika, że na drodze ewolucji osobnik będzie się uczył najlepszej strategii ewolucyjnej.

Jak można adaptować σ

Uwaga 1:

Teoretycznie powinno się zachować oceny mutacji takie jak dla parametrów obiektu. Jednak praktyka wykazuje, że ich łamanie może prowadzić do polepszenia.

Uwaga 2:

Mutacja σ powinna mieć charakter *multiplikatywny*, bo siła mutacji z definicji jest dodatnia.

Uwaga 3:

W oparciu o model sferyczny wykazano, że oczekiwane σ powinno się zmieniać o stały współczynnik.

Sposoby mutacji σ

Reguła „log-normal"

$$\tilde{\sigma}_I = \sigma_I exp(\zeta_I),$$

gdzie ζ_I jest losową wartością $\zeta_I = \tau \mathcal{N}_I(0,1)$, a τ — parametr uczenia. W lit. podano następujące wartości $\tau = 1/\sqrt{N}$ a w przestrzeni silnie multimodalnej $\tau = 1/\sqrt{2N}$

two-point rule

$$\tilde{\sigma}_I = \begin{cases} \sigma_I \cdot \alpha & if \quad u(0,1] \leq 1/2 \\ \sigma_I / \alpha & if \quad u(0,1] > 1/2 \end{cases},$$

gdzie $u(0,1]$ wartość losowa o rozkładzie jednostajnym, a α — parametr uczenia.

Sposoby mutacji σ cd.

Rozszerzona reguła „log-normal”

$$\tilde{\sigma}_I = exp(\tau_0\mathcal{N}_0(0,1)) \cdot (\sigma_1 exp(\tau\mathcal{N}_1(0,1)), \ldots, \sigma_N exp(\tau\mathcal{N}_N(0,1)))$$

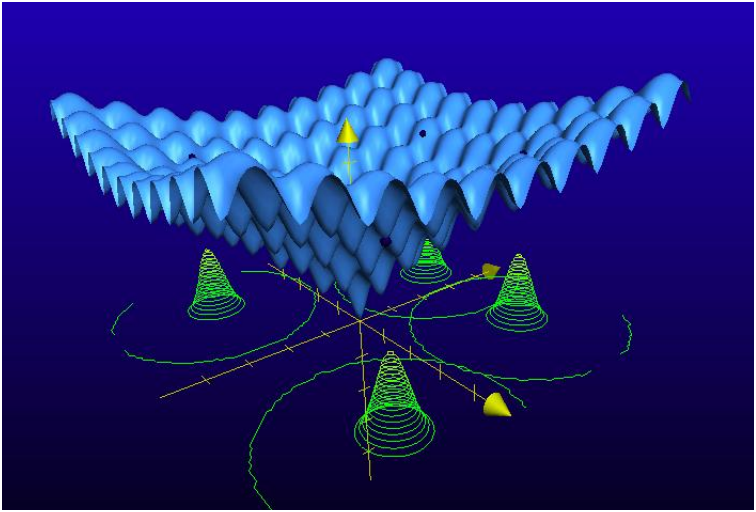
Główny mnożnik mutacji z parametrem τ_0 i adaptacyjne mutacje wzdłuż koordynat z parametrem τ.

Każdy parametr σ jest mutowany niezależnie a następnie cały wektor jest skalowany przez losowy współczynnik $exp(\tau_0\mathcal{N}_0(0,1))$. Rekomendowane współczynniki uczenia, to:

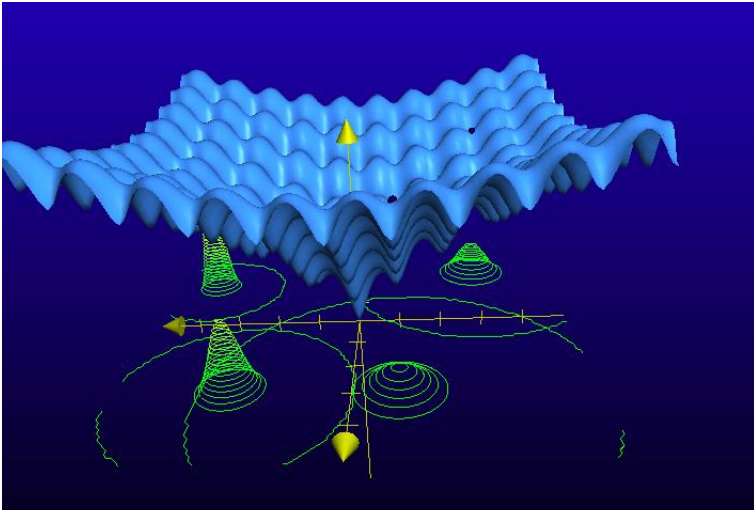
$$\tau_0 = \frac{c}{\sqrt{2N}} \quad and \quad \tau = \frac{c}{\sqrt{2\sqrt{N}}}$$

przy czym $c = 1$ dla (10,100)-ES

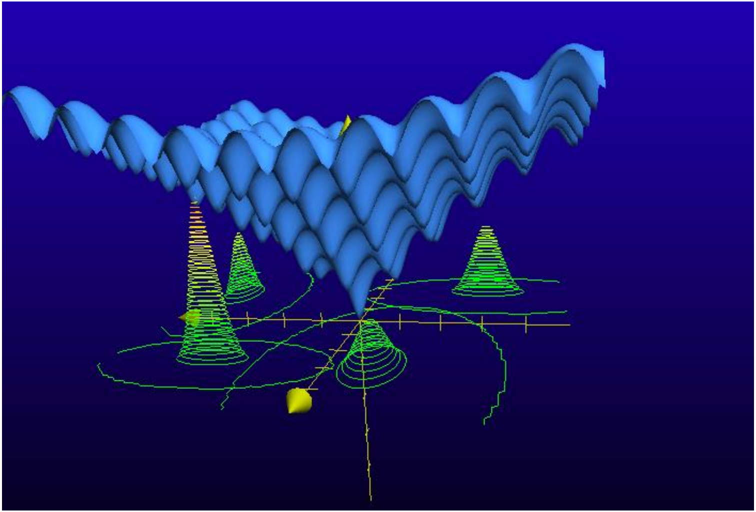
Bez mutacji σ



Mutacja σ



Mutacja σ_i



Rekombinacja parametrów strategii

Celem rekombinacji jest, tak jak w przypadku wektora **y**, ekstrakcja podobieństw pomiędzy rodzicami.

Dynamiczne zmiany parametrów strategii powodują, iż proces ewolucji staje się zaszumiony z silnymi wahaniami.

Jako że rekombinacja uśredniająca (μ/μ) (*intermediate*) ma właśnie cechy niwelowania wahań zalecane jest jej stosowanie do adaptacji parametrów endogennych.

Plan wykładu

Wykorzystana literatura (do samodzielnego studiowania)

- Thomas Weise
  Global Optimization Algorithms - Theory and Application.
  online e-book: http://www.it-weise.de/

- H.G. Beyer, H.P. Schwefel
  Evolution strategies. A comprehensive introduction.
  Natural Computing 1, p.3-52

- MARCELO DE BRITO
  GENETIC ARGONAUT
  http://geneticargonaut.blogspot.com/2006/03/evolutionary-computation-classics-vol.html