Równoważność algorytmów optymalizacji

Reguła „nie ma nic za darmo” (ang. *no free lunch theory*):

efektywność różnych typowych algorytmów szukania uśredniona po wszystkich możliwych problemach optymalizacyjnych jest taka sama

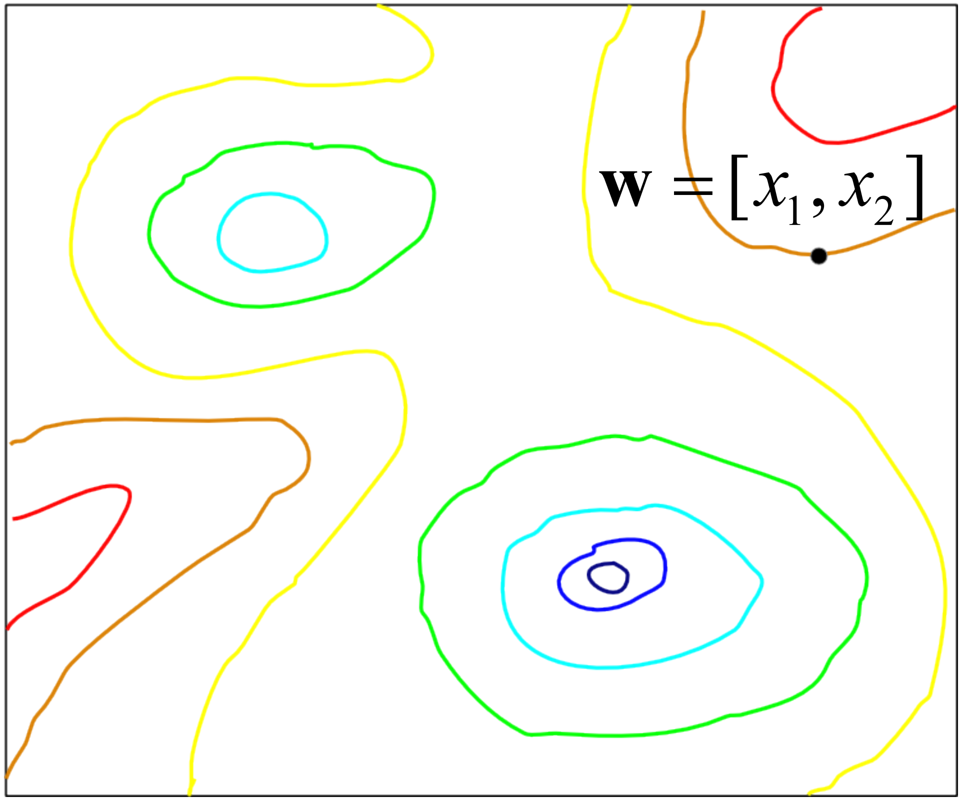
Typowe algorytmy szukania (optymalizacji):
- metoda enumeracyjna (wyliczeniowa),
- błądzenie przypadkowe,
- szukanie gradientowe,
- szukanie przypadkowe,
- symulowane wyżarzanie,
- algorytmy genetyczne,
- metody wyspecjalizowane (wykorzystujące szczegółową wiedzę o problemie).

Przykład: szukanie minimum funkcji

Szukanie enumeracyjne i błądzenie przypadkowe są nieefektywne w przypadku dużej liczby wymiarów.

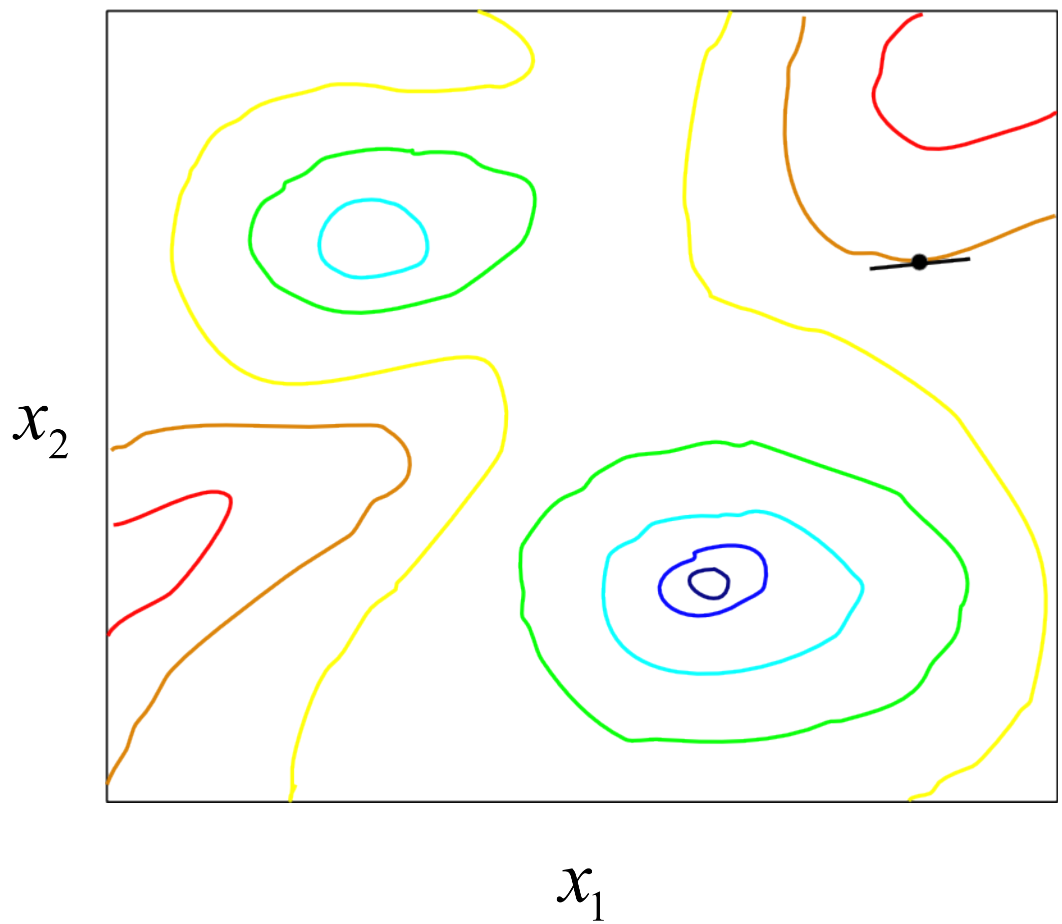
Przykład: szukanie minimum funkcji

Wybór punktu początkowego:



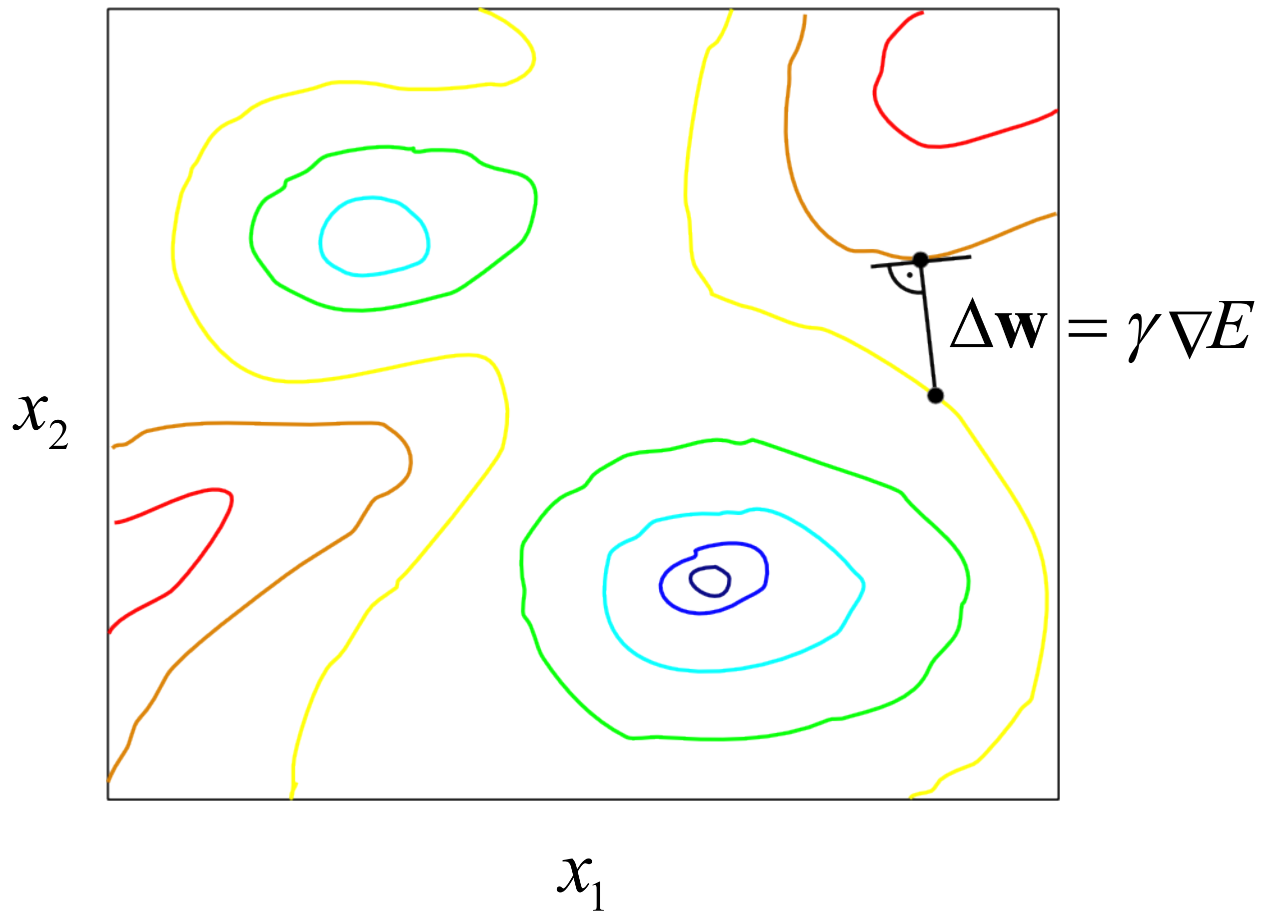
Przykład: szukanie minimum funkcji

Gradient: $\nabla E = [\frac{\partial E}{\partial x_1}, \frac{\partial E}{\partial x_2}]$ wyznacza kierunek największego spadku funkcji $E(x_1, x_2)$



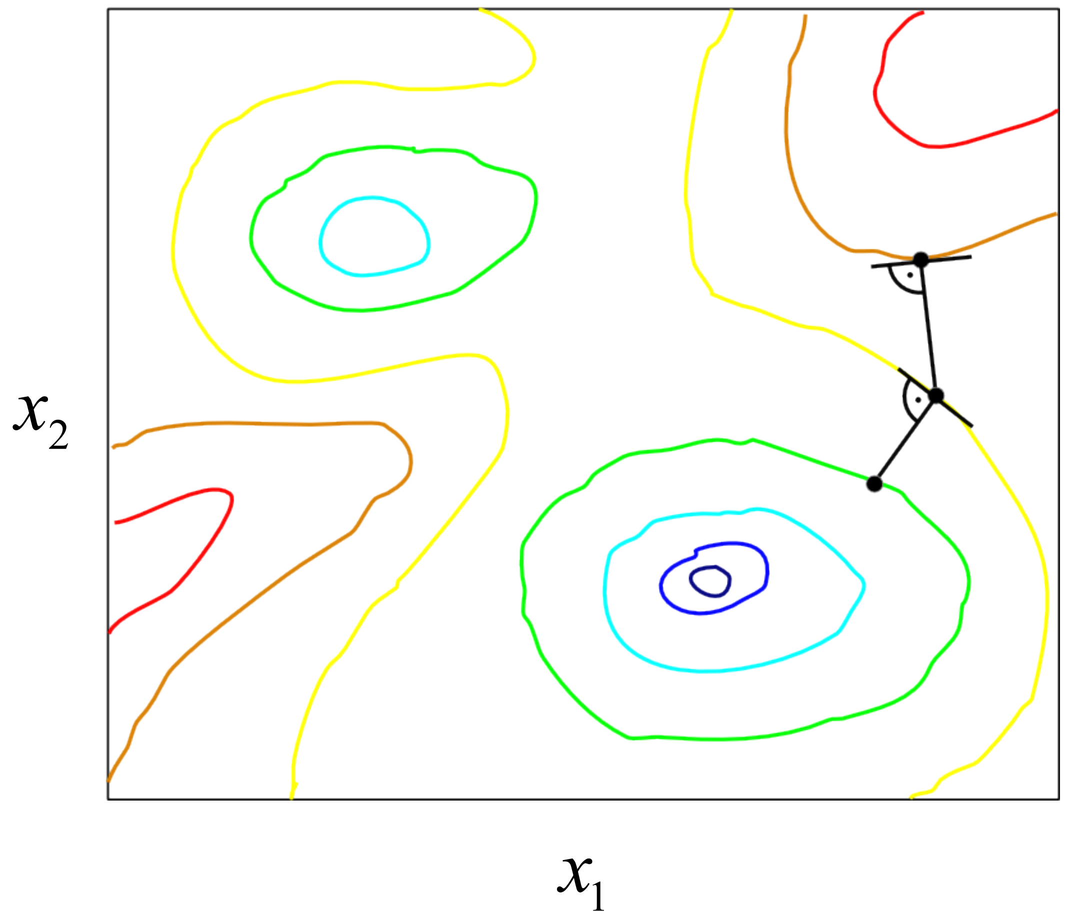
Przykład: szukanie minimum funkcji

Gradient: $\nabla E = [\frac{\partial E}{\partial x_1}, \frac{\partial E}{\partial x_2}]$ wyznacza kierunek największego spadku funkcji $E(x_1, x_2)$



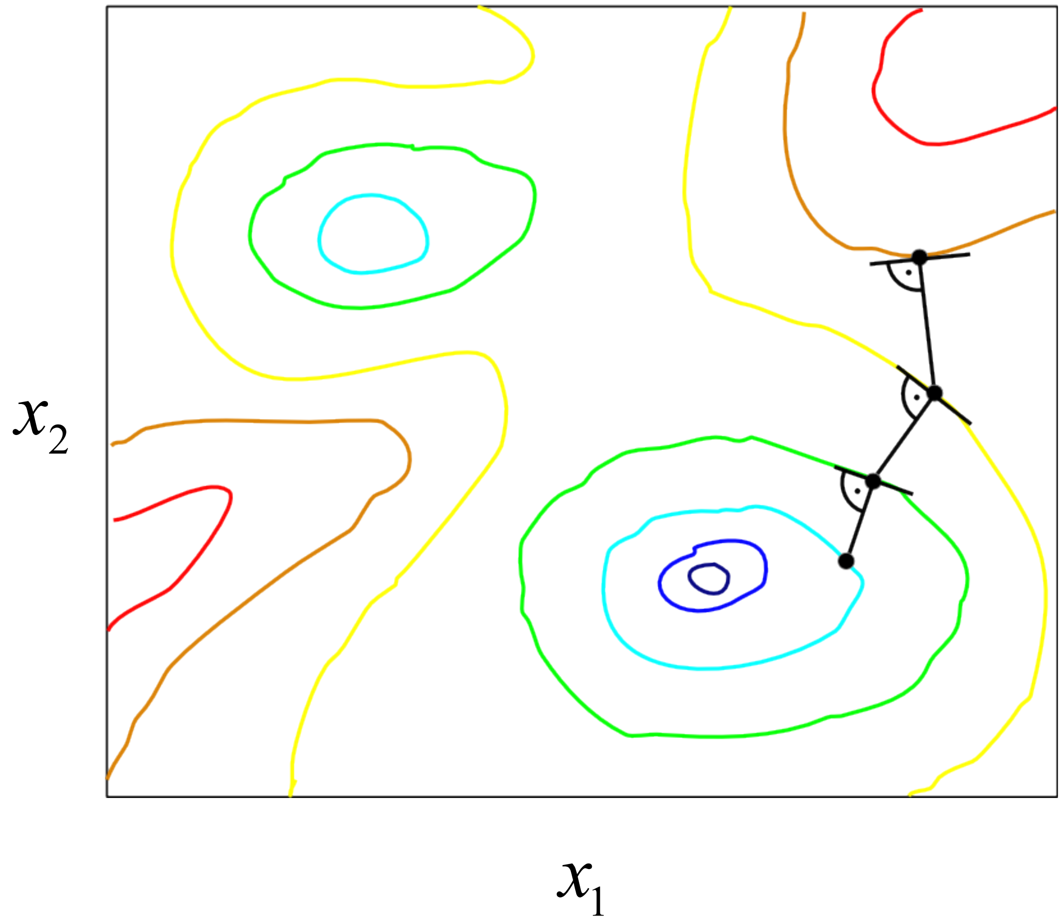
Przykład: szukanie minimum funkcji

Gradient: $\nabla E = [\frac{\partial E}{\partial x_1}, \frac{\partial E}{\partial x_2}]$ wyznacza kierunek największego spadku funkcji $E(x_1, x_2)$



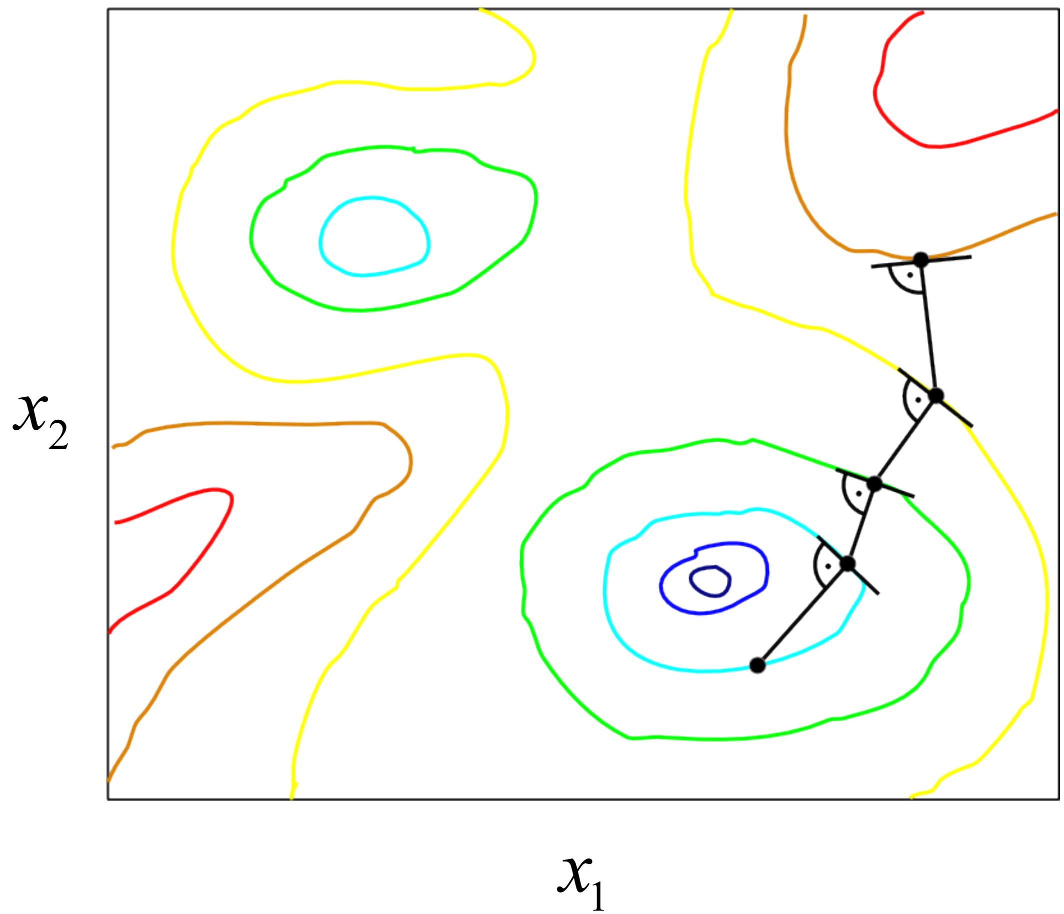
Przykład: szukanie minimum funkcji

Gradient: $\nabla E = [\frac{\partial E}{\partial x_1}, \frac{\partial E}{\partial x_2}]$ wyznacza kierunek największego spadku funkcji $E(x_1, x_2)$



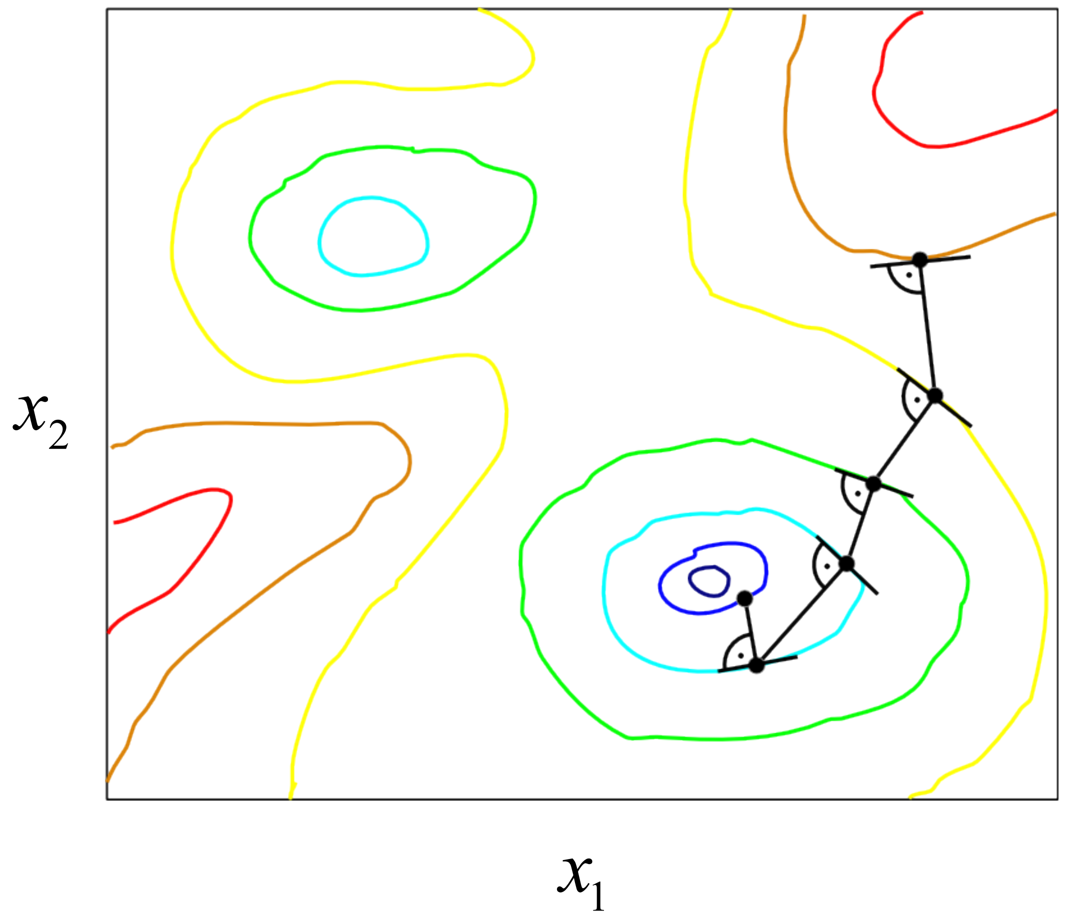
Przykład: szukanie minimum funkcji

Gradient: $\nabla E = [\frac{\partial E}{\partial x_1}, \frac{\partial E}{\partial x_2}]$ wyznacza kierunek największego spadku funkcji $E(x_1, x_2)$



Przykład: szukanie minimum funkcji

Gradient: $\nabla E = [\frac{\partial E}{\partial x_1}, \frac{\partial E}{\partial x_2}]$ wyznacza kierunek największego spadku funkcji $E(x_1, x_2)$



Przykład: szukanie minimum funkcji

Wady metody gradientowej:

- konieczna jest różniczkowalność funkcji we wszystkich wymiarach,
- wrażliwość na niewielkie ,,nierówności" powierzchni, rozwiązaniem może być metoda z momentem: $\Delta\mathbf{w}_t = \alpha\Delta\mathbf{w}_{t-1} + (1-\alpha)\gamma\,\nabla E_t$, gdzie α - współczynnik momentu,
- zbieżność do minimum lokalnego.

Przykład: szukanie minimum funkcji

Szukanie przypadkowe: losowanie nowego rozwiązania w otoczeniu rozw. dotychczasowego; akceptacja tylko rozwiązania lepszego.

Przykład: szukanie minimum funkcji

Szukanie przypadkowe: losowanie nowego rozwiązania w otoczeniu rozw. aktualnego; akceptacja tylko rozwiązania lepszego.

Przykład: szukanie minimum funkcji

Szukanie przypadkowe: losowanie nowego rozwiązania w otoczeniu rozw. aktualnego; akceptacja tylko rozwiązania lepszego.

Przykład: szukanie minimum funkcji

Szukanie przypadkowe: losowanie nowego rozwiązania w otoczeniu rozw. aktualnego; akceptacja tylko rozwiązania lepszego.

Przykład: szukanie minimum funkcji

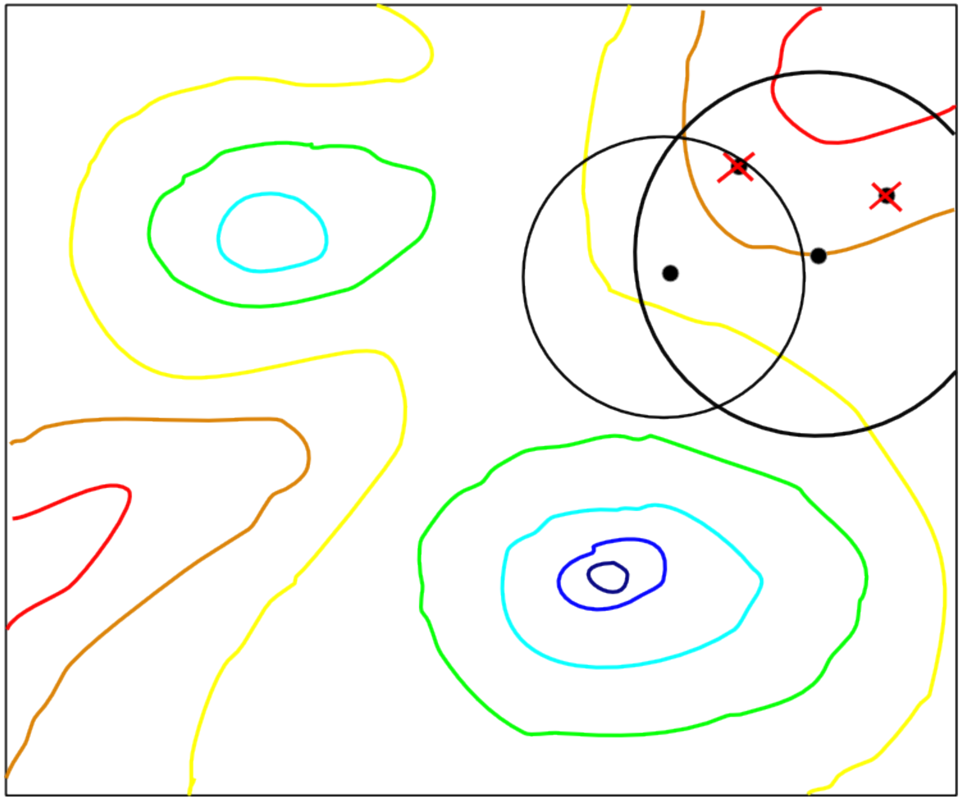
Szukanie przypadkowe: losowanie nowego rozwiązania w otoczeniu rozw. aktualnego; akceptacja tylko rozwiązania lepszego.

Przykład: szukanie minimum funkcji

Szukanie przypadkowe: losowanie nowego rozwiązania w otoczeniu rozw. aktualnego; akceptacja tylko rozwiązania lepszego.

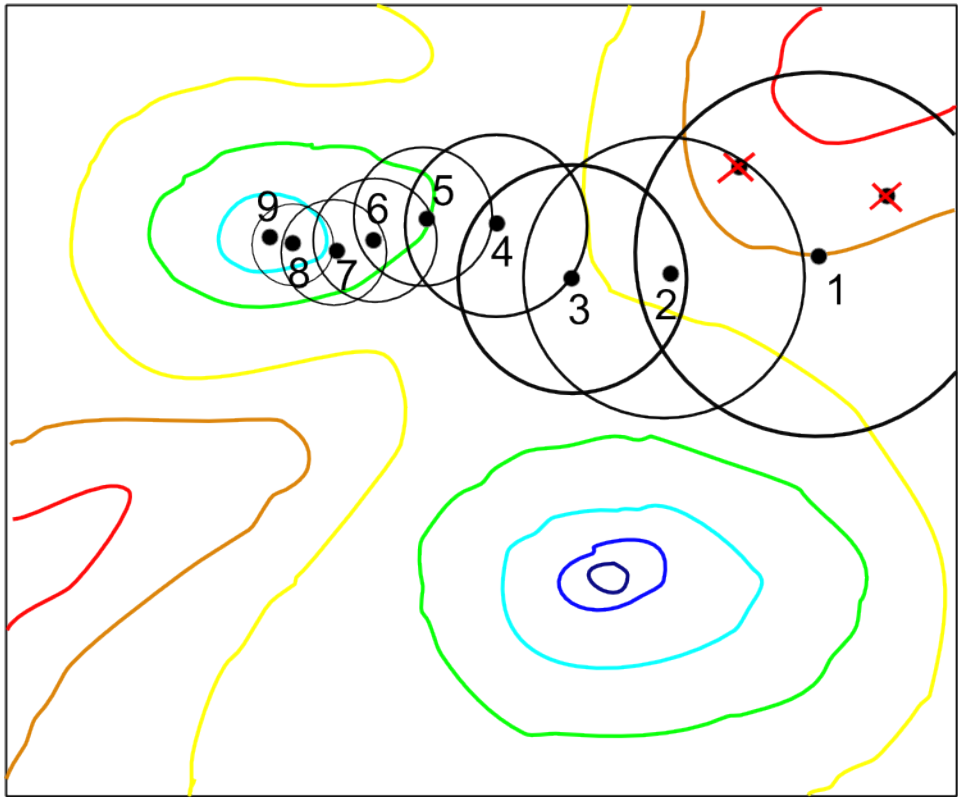
Przykład: szukanie minimum funkcji

Szukanie przypadkowe: losowanie nowego rozwiązania w otoczeniu rozw. aktualnego; akceptacja tylko rozwiązania lepszego.



Przykład: szukanie minimum funkcji

Szukanie przypadkowe: losowanie nowego rozwiązania w otoczeniu rozw. aktualnego; akceptacja tylko rozwiązania lepszego.



Wada: zbieżność do minimum lokalnego.

Symulowane wyżarzanie

Prawdopodobieństwo akceptacji rozwiązania gorszego niż dotychczasowe może być niezerowe:

$$h(\Delta E, T) = \frac{1}{1 + \exp\left(\Delta E / cT\right)}$$

gdzie:

$\Delta E = E(\mathbf{w}_t) - E(\mathbf{w}_{t-1})$ - zmiana wartości rozwiązania,

T - temperatura, zwykle jest funkcją malejącą,

c - stała pozwalająca na zróżnicowanie wpływu temperatury na prawdopodobieństwo akceptacji i promień obszaru z którego losowane jest nowe rozwiązanie.

Symulowane wyżarzanie

Prawdopodobieństwo akceptacji rozwiązania:

$$h(\Delta E, T) = \frac{1}{1 + \exp\left(\Delta E / cT\right)}$$

$T \to 0$

$$\Delta E > 0 \quad \left(\Delta E / cT\right) \to \infty \quad h \to 0$$

$$\Delta E < 0 \quad \left(\Delta E / cT\right) \to -\infty \quad h \to 1$$

(szukanie przypadkowe)

$T \to \infty$

$$\left(\Delta E / cT\right) \to 0 \quad h \to \frac{1}{2}$$

(błądzenie przypadkowe)

Algorytmy genetyczne - plan wykładu

- Standardowy algorytm genetyczny
- Czynności związane z realizacją AG
- Techniki poprawiające efektywność algorytmów genetycznych
- Programowanie genetyczne
- Genetyczny dobór wag SSN

Algorytmy genetyczne - literatura

- John Holland, ***Adaptation in natural and artificial systems***, The University of Michigan Press, 1975
- David E. Goldberg, ***Algorytmy genetyczne i ich zastosowania***, WNT, Warszawa 1995
- Jarosław Arabas, ***Wykłady z algorytmów ewolucyjnych***, WNT, Warszawa 2001

Algorytmy genetyczne

Algorytmy genetyczne – algorytmy poszukiwania oparte na mechanizmach doboru naturalnego oraz łączenia cech rozwiązań

Cele badań nad algorytmami genetycznymi:

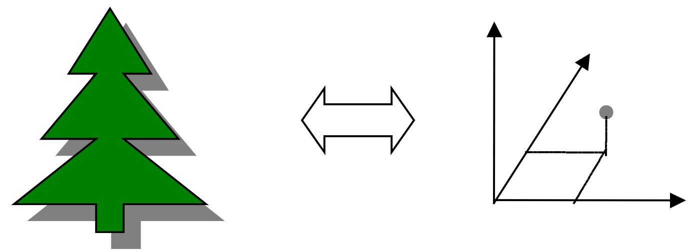
1. wyjaśnienie procesów adaptacyjnych występujących w przyrodzie,
2. zastosowanie w zadaniach optymalizacji i uczenia.

Terminy genetyczne i ich odpowiedniki w sztucznych systemach genetycznych

fenotyp	rozwiązanie w postaci zbioru parametrów - **x**
genotyp	zakodowana postać rozwiązania (fenotypu)
chromosom	ciąg kodowy składający się z genów - cech
allel	wariant (wartość) cechy
locus	pozycja genu w chromosomie
przystosowanie	funkcja $f(\mathbf{x})$ przypisująca każdemu rozwiązaniu liczbę rzeczywistą odzwierciedlającą jego wartość

Terminy genetyczne - schemat

fenotyp:

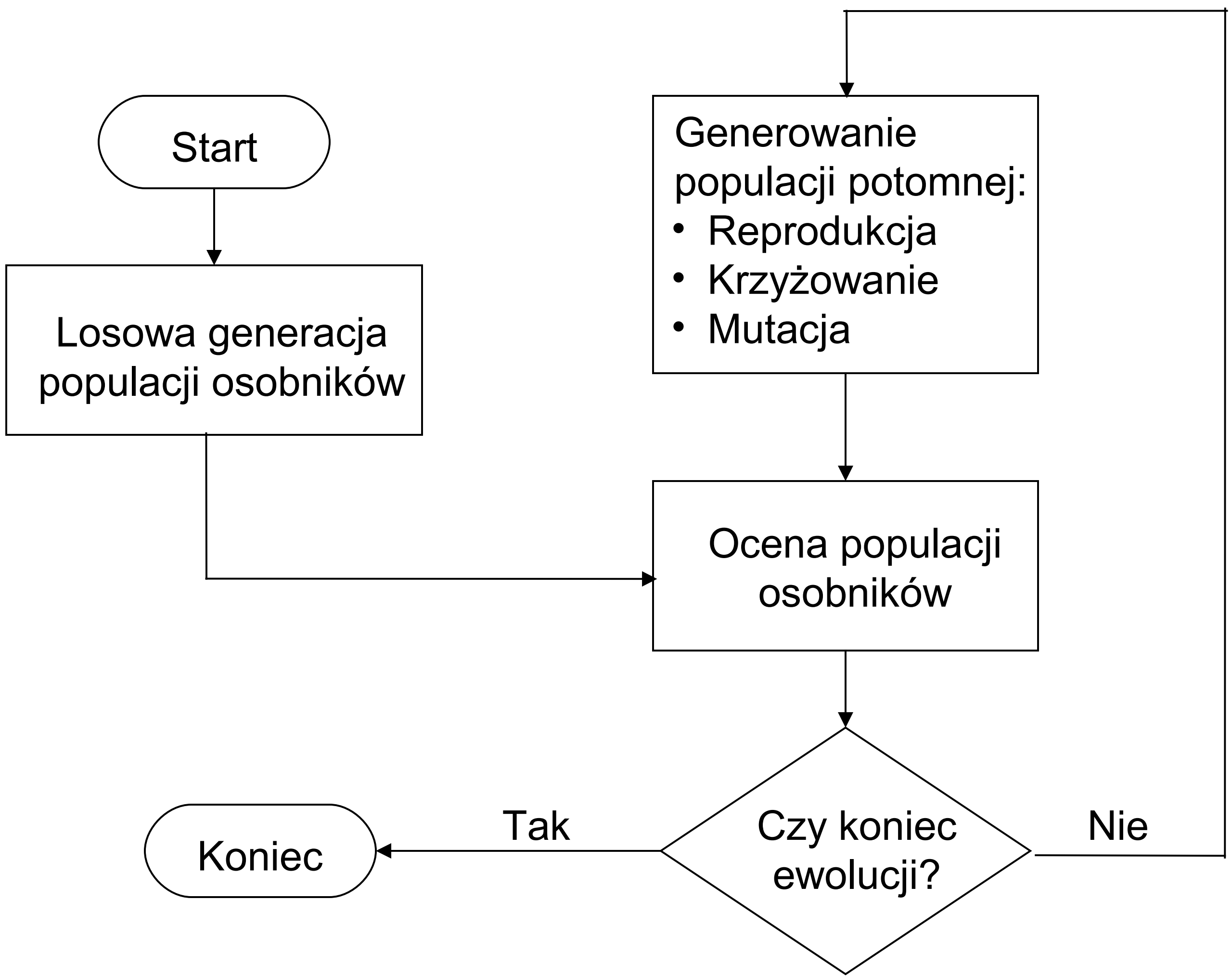


genotyp:

$$1\ \ 0\ \ \underbrace{\boxed{1}}_{\text{gen } x}\ \ 0\ \ 0\ \ 1\ \ 0$$

allel(gen x) = 1
zbiór alleli (gen x) = {0,1}
locus(gen x) = 3

Standardowy algorytm genetyczny – ogólny schemat



Funkcje podstawowych operatorów genetycznych

Reprodukcja – wybór najlepiej przystosowanych osobników (rozwiązań) do następnego pokolenia

Krzyżowanie – szukanie rozwiązań zawierających cechy wielu dobrych rozwiązań

Mutacja – dostarczanie nowego materiału genetycznego

Typy reprodukcji

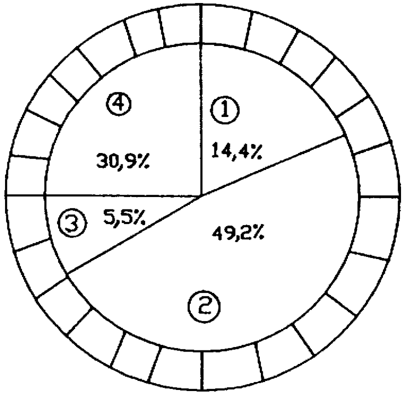
- Ruletkowa – proporcjonalna do wartości funkcji przystosowania - konieczność skalowania
- Rangowa – zależna od rangi – numeru na liście posortowanej względem przystosowania
- Turniejowa – wielokrotny wybór najlepszego osobnika z losowo wybieranej podpopulacji aż do skompletowania populacji potomnej

Reprodukcja ruletkowa

Przykład: optymalizacja funkcji $f(x) = x^2$

Nr	Ciąg kodowy	Przystosowanie	% całości
1	01101	169	14,4
2	11000	576	49,2
3	01000	64	5,5
4	10011	361	30,9
Łącznie		1170	100,0

Koło ruletki:



Reprodukcja rangowa

Metody wyznaczania prawdopodobieństwa reprodukcji:

- Liniowa: $p(X) = a + k(1 - \frac{r(X)}{r_{\max}})$
- Potęgowa: $p(X) = a + k(r_{\max} - r(X))^b$

gdzie:
r(X) – ranga rozwiązania X,
$r_{\max}$ – ranga maksymalna,
a,b,k – stałe spełniające warunki:

$$\sum_{i=1}^{N} p(X_i) = 1,$$

$$0 \leq p(X) \leq 1,$$

$$r(X) \geq r(Y) \Rightarrow p(X) \leq p(Y)$$

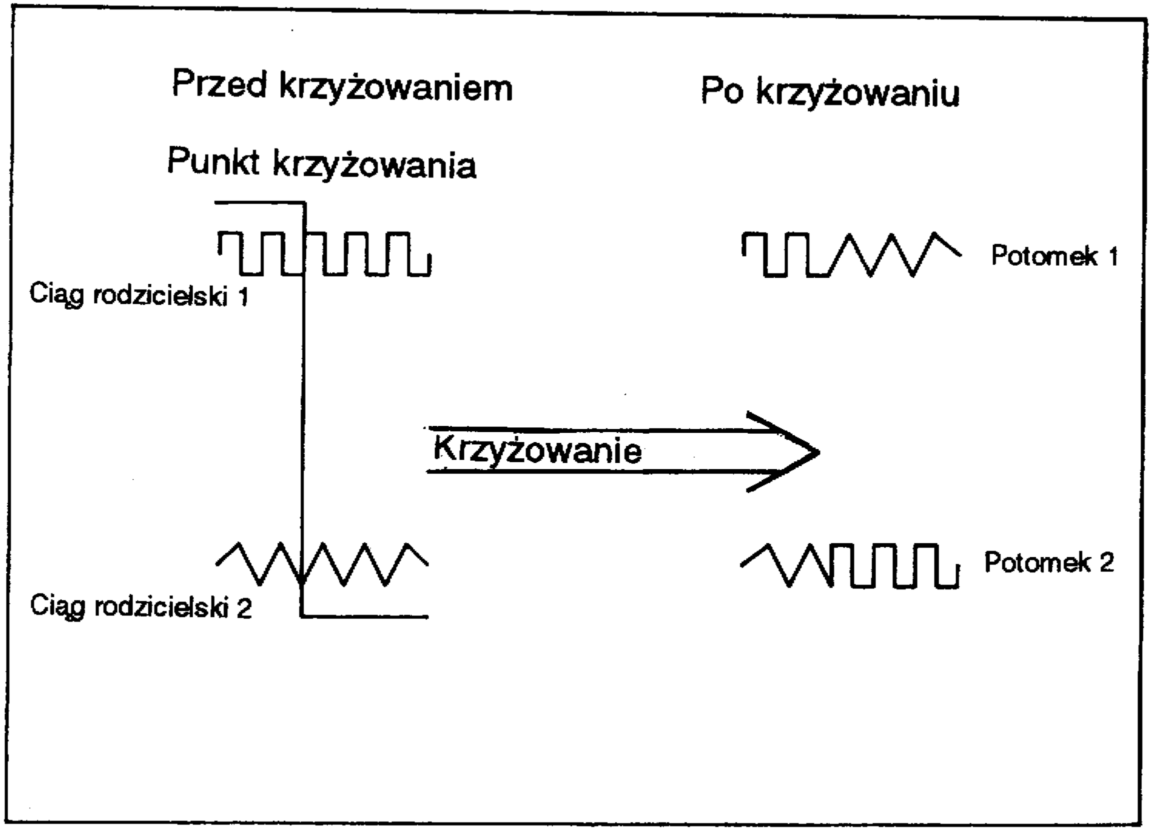
Reproducja rangowa - przykład

Wyznaczanie prawdopodobieństwa reprodukcji metodą liniową dla funkcji przystosowania $f(x) = x^2$

$$p(X) = a + k(1 - \frac{r(X)}{r_{\max}})$$

Nr	Ciąg kodowy	Przystosowanie	Ranga *r(X)*	$1 - \frac{r(X)}{r_{\max}}$	*p(X)* dla k=2/3 i a=0
1	01101	169	3	0,25	0,17
2	11000	576	1	0,75	0,50
3	01000	64	4	0	0
4	10011	361	2	0,50	0,33
Łącznie		1170		1,5	1,0

Krzyżowanie proste - schemat



Symulacja algorytmu genetycznego optymalizującego funkcję $f(x) = x^2$

Ciągi kodowe												
Nr ciągu	Populacja początkowa (wygenerowana losowo)	Wartość x (Liczba całkowita)	$f(x)$ x^2	pselect$_i$ $\frac{f_i}{\sum f}$	Oczekiwana liczba kopii $\frac{f_i}{\bar{f}}$	Liczba kopii wylosowanych (wg reguły ruletki)	Pula rodzicielska po reprodukcji (zaznaczono punkty krzyżowania)	Partner (wybrany losowo)	Punkt krzyżowania (wybrany losowo)	Nowa populacja	Wartość x	$f(x)$ x^2
1	01101	13	169	0,14	0,58	1	0 1 1 0 | 1	2	4	0 1 1 0 0	12	144
2	11000	24	576	0,49	1,97	2	1 1 0 0 | 0	1	4	1 1 0 0 1	25	625
3	01000	8	64	0,06	0,22	0	1 1 | 0 0 0	4	2	1 1 0 1 1	27	729
4	10011	19	361	0,31	1,23	1	1 0 | 0 1 1	3	2	1 0 0 0 0	16	256
Suma			1170	1,00	4,00	4,0						1754
Średnia			293	0,25	1,00	1,0						439
Maksimum			576	0,49	1,97	2,0						729

Czynności związane z realizacją algorytmu genetycznego

- Wybór metody kodowania - reprezentacji fenotypu (rozwiązania)
- Wybór operatorów genetycznych w zależności od problemu i przyjętego sposobu kodowania
- Dobór wartości parametrów ewolucji
- Dobór funkcji oceny

Metody kodowania

Ze względu na typ wartości genu - allelu:

- Binarne np. 1000110
- Całkowitoliczbowe
- O wartościach rzeczywistych (fenotypowe)

Kodowanie fenotypowe

Wybrane operatory fenotypowe:

- mutacja fenotypowa normalna:

$$x_i' = x_i + N(0,\delta)$$

- krzyżowanie fenotypowe:

$$x_i' = x_i^1 + \gamma \cdot N(0,\delta) \cdot (x_i^2 - x_i^1), \quad i = 1,\ldots N$$

$\mathbf{x} = (x_1, x_2, \ldots, x_N)$ - wektor parametrów fenotypu

Metody kodowania

Ze względu na sposób reprezentowania cech:

- Klasyczne - każda cecha fenotypu jest kodowana przez wartość odpowiedniego genu niezależnie od jego umiejscowienia
- Permutacyjne - cechy kodowane są przez pozycje genów (*locus*)
- Mieszane - cechy kodowane zarówno przez pozycje jak i umiejscowienie genów

Problem komiwojażera – przykład kodowania permutacyjnego

Założenia:

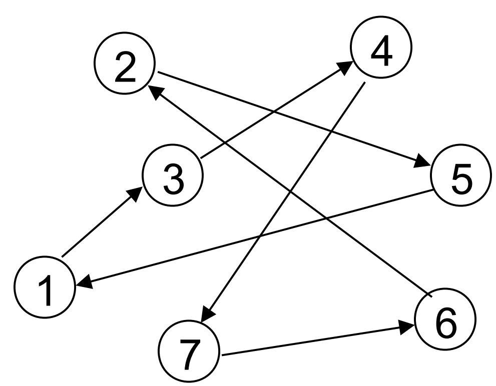
- każde miasto jest przypisane do jednego z genów
- o kolejności odwiedzin każdego z miast decyduje jego umiejscowienie w ciągu kodowym

Przykład:

ciąg kodowy:

2	5	1	3	4	7	6

rozwiązanie:



Metody kodowania - cd

Ze względu na strukturę genotypu:

- Za pomocą ciągów
- Za pomocą innych struktur np. drzew, grafów, sieci

Dobór funkcji oceny

Zasady:

- Oceny powinny być nieujemne i tym większe im większa jakość rozwiązań (ze względu na znane metody reprodukcji).
- Rozwiązanie w dowolnie małym stopniu lepsze powinno otrzymać większą ocenę (by możliwe było ciągłe ulepszanie rozwiązań).

Zadania dodatkowe:

- Dobór współczynnika selekcji, tak by osiągnąć kompromis pomiędzy różnorodnością populacji a ukierunkowaniem poszukiwań.
- Odpowiednie łączenie ocen cząstkowych w optymalizacji wielokryterialnej.

Dobór funkcji oceny

Łączenie ocen cząstkowych:

- Prosta suma: $k = k_1 + k_2 + ... + k_N$
- Z uwzględnieniem ryzyk/kosztów: $k = w_1 k_1 + w_2 k_2 + ... + w_N k_N$
- Z jednym jak najlepszym podrozwiązaniem: $k = \max(k_1, k_2, ..., k_N)$ lub $k = \max(w_1 k_1, w_2 k_2, ..., w_N k_N)$
- Ze wszystkimi dobrymi podrozwiązaniami: $k = \min(k_1, k_2, ..., k_N)$ lub $k = \min(w_1 k_1, w_2 k_2, ..., w_N k_N)$
- Uogólnienie:

$$k = f_p(k_1, k_2, \ldots, k_N) = (k_1^p + k_2^p + \ldots + k_N^p)^{\frac{1}{p}}$$

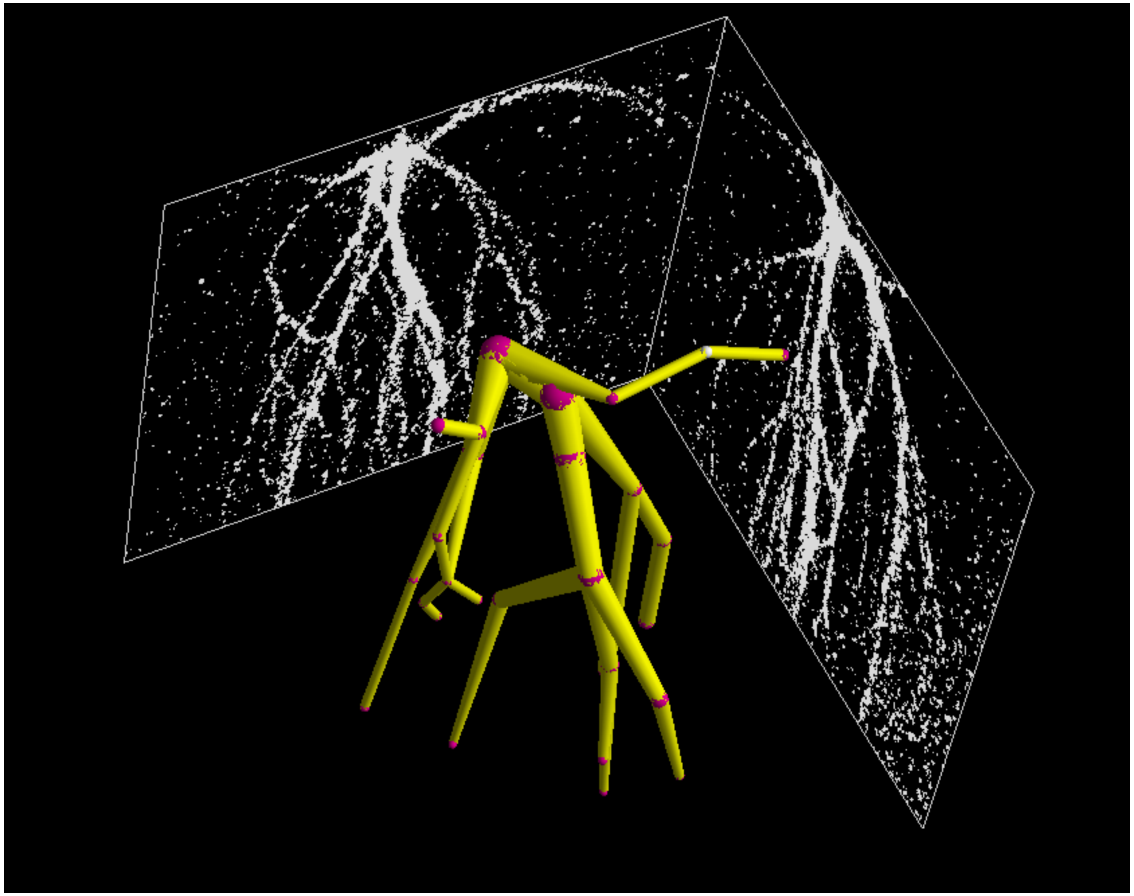
$$p \to \infty \quad, f_p = \max$$

$$p \to 0, f_p = const$$

$$p \to -\infty \ , f_p = \min$$

Dobór funkcji oceny - przykład

Modelowanie wyładowań elektrycznych (obiektów trójwymiarowych) na podstawie pary zdjęć.



Dobór funkcji oceny - przykład

Modelowanie wyładowań elektrycznych (obiektów trójwymiarowych) na podstawie pary zdjęć - umożliwienie ciągłego ulepszania rozwiązań:

- utworzenie macierzy minimalnych odległości od jasnych punktów na zdjęciach,
- uwzględnienie odległości w funkcji oceny rozwiązań (rozwiązanie otrzymuje tym wyższą ocenę, im poszczególne punkty na rzutach modelu znajdują się bliżej jasnych punktów na zdjęciach).

Model elitarny

Najlepsze osobniki w populacji przechodzą do następnego pokolenia bez jakichkolwiek zmian.

zaleta - zachowanie najlepszych znanych rozwiązań
wada - nadmierne skupienie populacji w obszarach wybranych rozwiązań

Zalety i wady algorytmów genetycznych

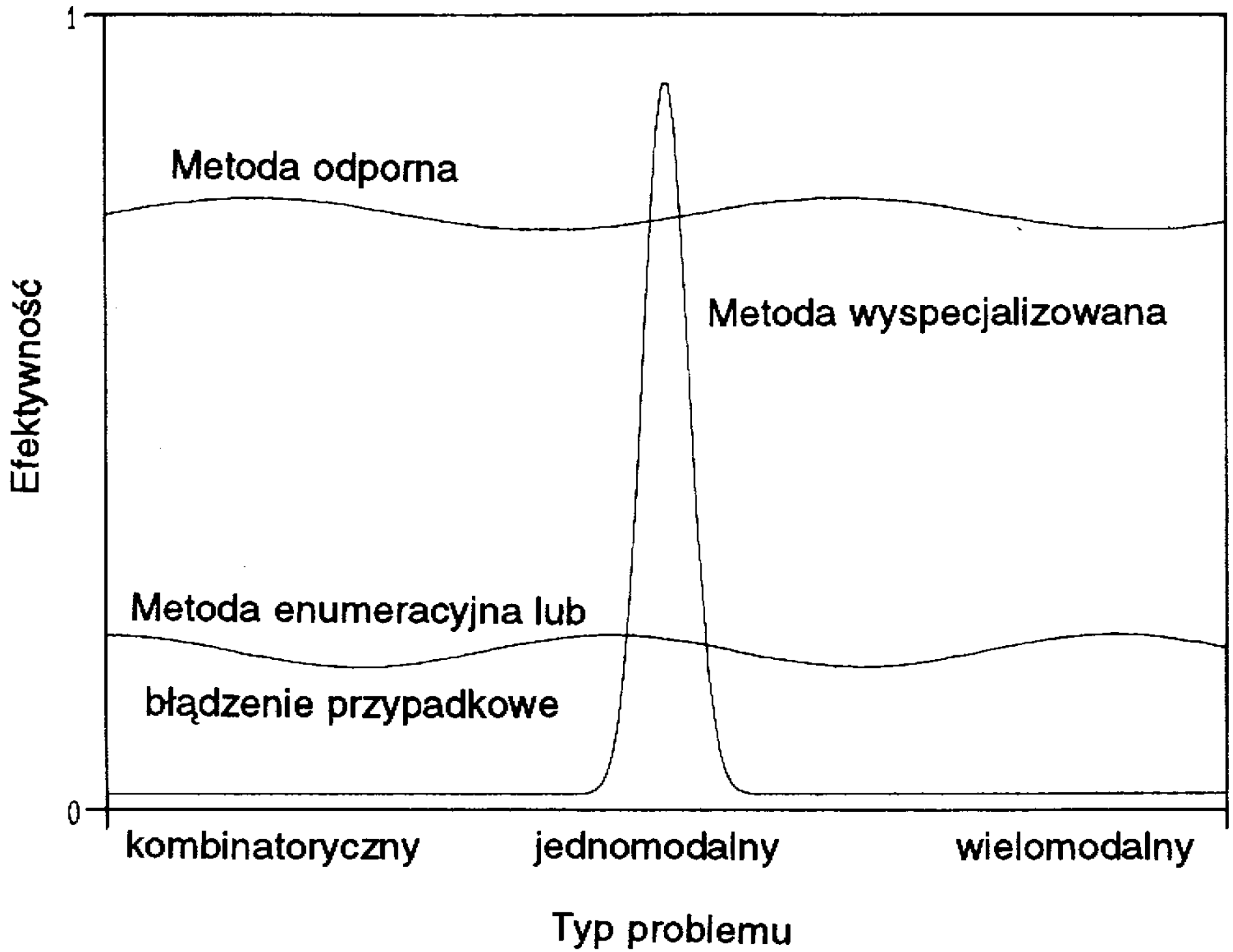
Zalety:

- Odporność - unikanie ekstremów lokalnych, prawdopodobieństwo znalezienia dobrych rozwiązań jest w dużym stopniu niezależne od wyboru punktów początkowych
- Wydajność – duża liczba przetwarzanych schematów - ok. m^3, gdzie m - liczba osobników w populacji
- Łatwość zastosowania w niemal każdym zadaniu optymalizacji

Wady:

- Brak gwarancji zbieżności do optymalnego rozwiązania

Metody poszukiwań - porównanie



Różnice pomiędzy algorytmami genetycznymi a tradycyjnymi metodami szukania

- Algorytmy genetyczne przetwarzają zakodowaną postać parametrów zadania (ciąg kodowy) a nie same parametry
- Poszukiwania prowadzone są w obrębie całej populacji rozwiązań (osobników) a nie pojedynczego rozwiązania
- Wykorzystywana jest tylko funkcja celu (uczenie z krytykiem) bez żadnej dodatkowej informacji naprowadzającej np. pochodnej funkcji celu
- Stosowane są probabilistyczne a nie deterministyczne reguły wyboru

Metody likwidowania ograniczeń standardowego algorytmu genetycznego

Ograniczenia	Metody przezwyciężania ograniczeń
niestabilność, brak zbieżności przy braku dobrych schematów	rekonfiguracja (zmiana położenia genów bez zmiany rozwiązania)
przedwczesna zbieżność	zwiększenie mutacji model ze ściskiem, mechanizm preselekcji, metody niszowe
niestacjonarność środowiska	diploidalność, poliploidalność, w pewnym stopniu metody zachowania różnorodności populacji (zapobiegania przedwczesnej zbieżności)

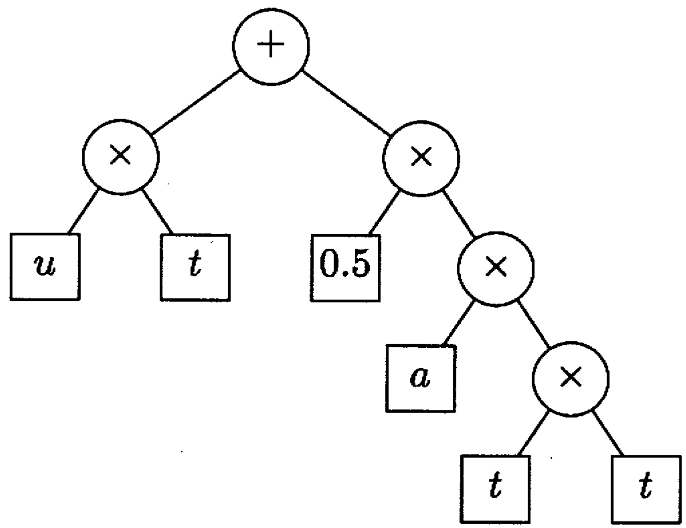
Programowanie genetyczne

GP - *(genetic programming)*

Projektowanie GP:

- wybór alfabetu (np. funkcje logiczne, arytmetyczne, statystyczne, instrukcje i bloki danych)
- wybór struktury (np. drzewo, graf, automat skończony)
- wybór operatorów krzyżowania i mutacji
- wybór operatora enkapsulacji

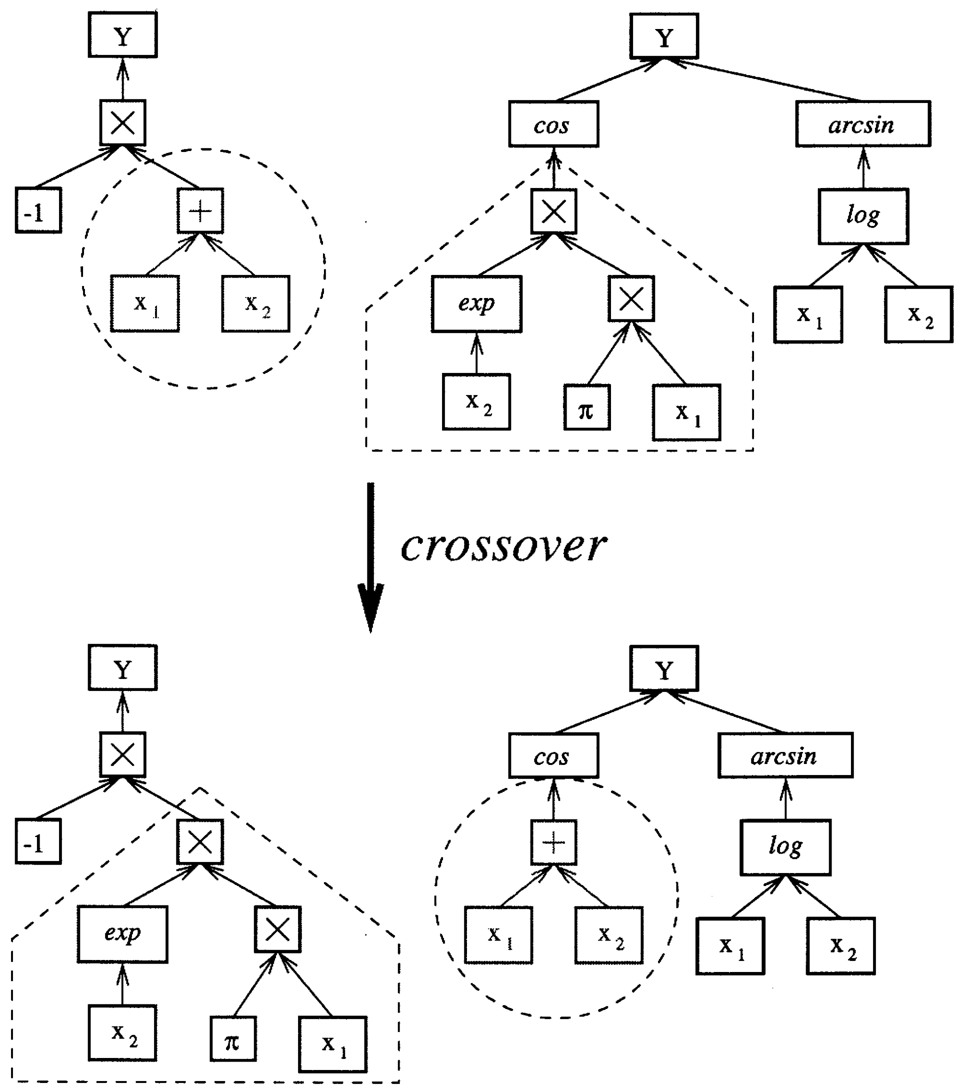
Programowanie genetyczne - kodowanie funkcji w postaci symbolicznej



$$+(\times(u,t),\times(0.5,\times(a,\times(t,t))))$$

An example of a genetic program to compute $s = ut + \frac{1}{2}at^2$.

Programowanie genetyczne - krzyżowanie



Programowanie genetyczne - enkapsulacja