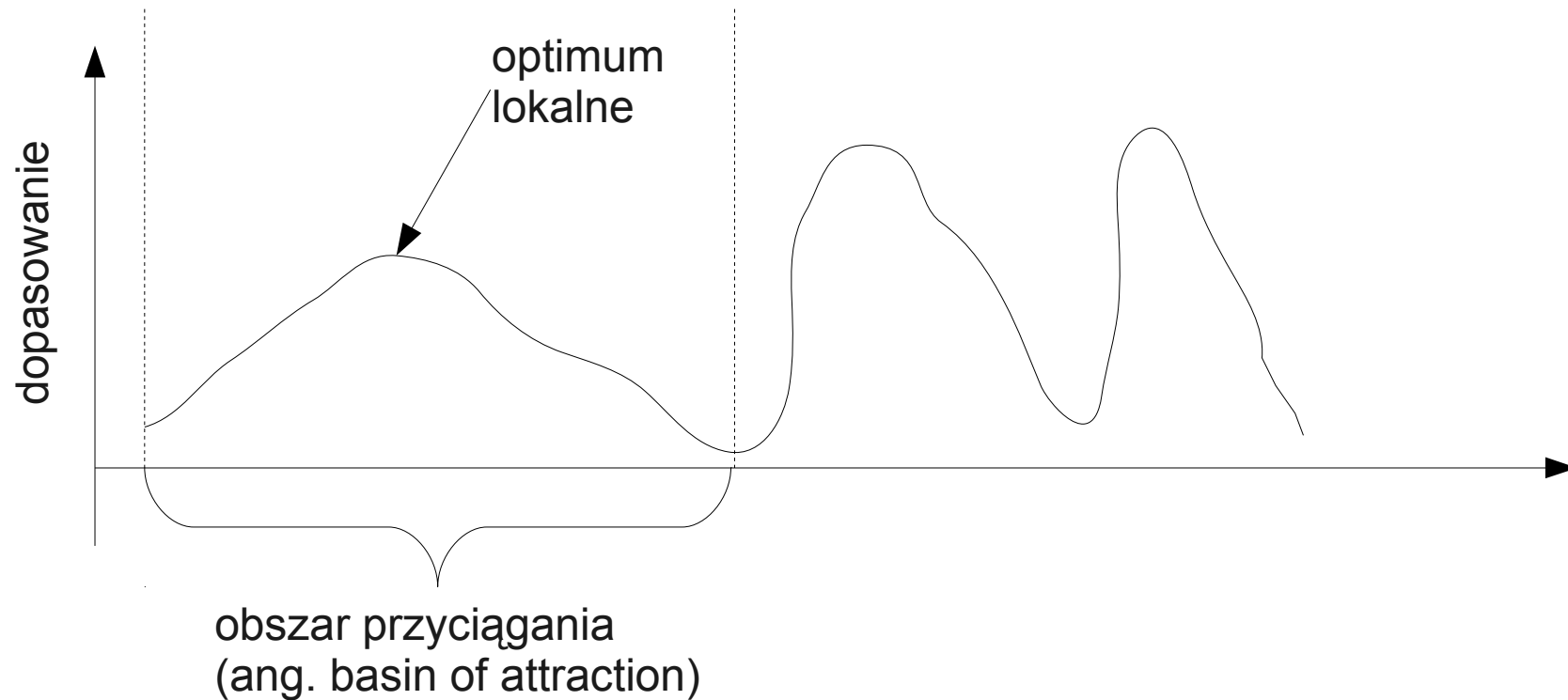


Problemy multimodalne, rozdzielone populacje
oraz optymalizacja wielokryterialna

Wprowadzenie

- We wszystkich algorytmach ewolucyjnych omawianych do tej pory, wszystkie osobniki były elementami jednej wspólnej populacji.
- Na ewolucje biologiczną wpływ ma jeden istotny czynnik: fizyczna przestrzeń na której ta ewolucja się odbywa.
 - Kwiaty z mojego ogrodu prawie na pewno nie zostaną skrzyżowane z kwiatami z pewnego ogrodu w Australii.
- Osobnik ma szansę być skrzyżowanym (i porównanym w procesach selekcji oraz sukcesji) tylko z osobnikami żyjącymi w lokalnym sąsiedztwie.
 - To zjawisko może promować **różnorodność** (ang. diversity) w globalnej populacji, a lokalne grupy mogą się adaptować do lokalnych warunków, co zostało już opisane przez Darwina w słynnej książce *The Origin of Species*.
- Pomysły takie jak podział populacji na mniejsze podpopulacje, z rzadka komunikujące się oraz wprowadzenie ograniczeń w krzyżowaniu (nie każdy osobnik może krzyżować się z każdym) zostały wykorzystane w dyscyplinie obliczeń ewolucyjnych.

Wielomodalna funkcja dopasowania

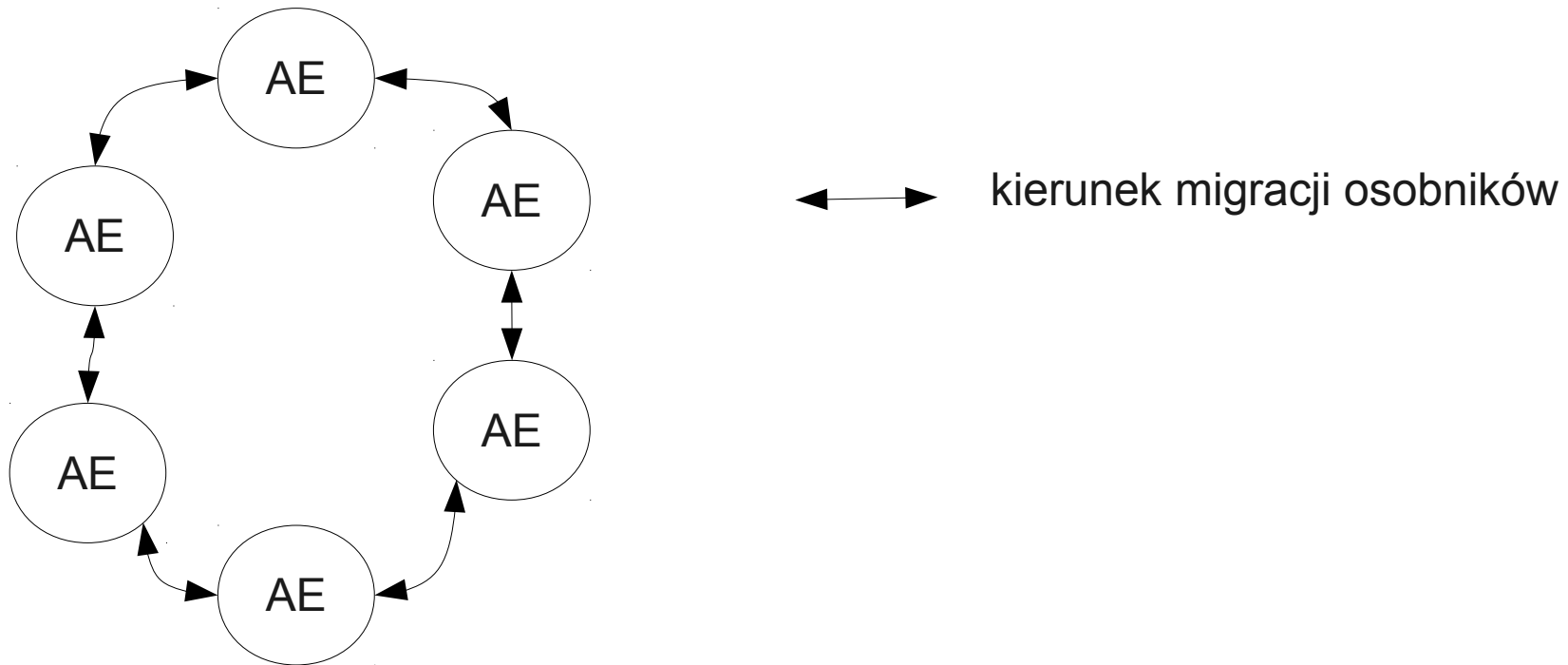


- Nisza (ang. niche) - obszar o wysokim dopasowaniu nie połączony z innymi tego typu obszarami.
- Dryf genetyczny (genetic drift) - w algorytmie z jedną globalną populacją o skończonym rozmiarze, w którym rekombinacja może zachodzić pomiędzy dowolną parą rodziców, a selekcja operuje na całej populacji, nastąpi zbieżność wokół jednej niszy.
- Możemy chcieć zidentyfikować wiele optimumów:
 - aby zwiększyć szansę na znalezienie optimum globalnego.
 - aby pozostawić decyzję o wyborze finalnego rozwiązania człowiekowi.

Metody pozwalające na otrzymanie zbioru różnorodnych rozwiązań problemu

- Metody niejawne, nie wymagające wymuszania różnorodności w populacji
 - Uruchom algorytm wiele razy, startując za każdym razem z innego stanu początkowego. Problemатyczne, gdy jedno z optimów ma b. duży obszar przyciągania.
 - Uruchom kilka algorytmów genetycznych współbieżnie (być może równoległe, jeżeli mamy hardware) pozwól im co jakiś czas dzielić się informacjami. Technika zwana modelem **wyspowym** (ang. island).
 - Wprowadź przestrzenny podział pojedynczej populacji, w którym każdy osobnik zna pewną liczbę swoich sąsiadów a selekcja, krzyżowanie, mutacja i sukcesja ograniczone są do sąsiedztwa osobnika (model **dyfuzyjny**).
- Metody jawne, wymuszające podział populacji na niesze
 - Współdzielenie dopasowania (ang. fitness sharing) przez podobne osobniki
 - Ścisk (ang. crowding)

Model wyspowy

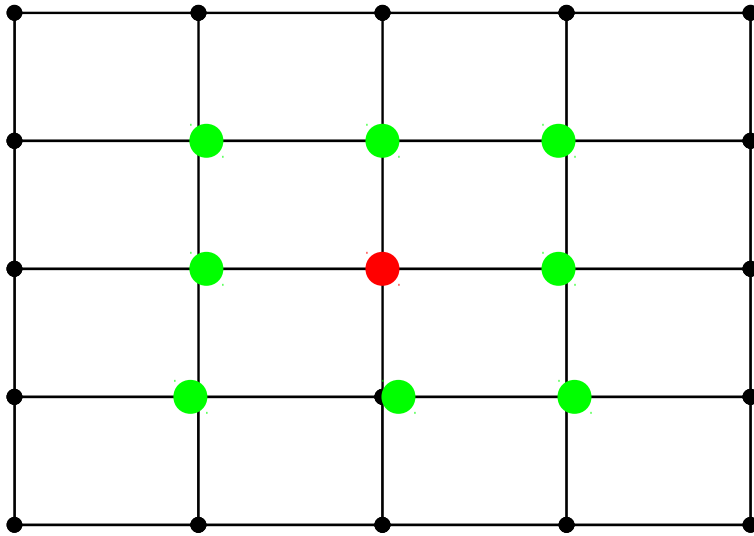


- Model zaprojektowany z myślą o maszynach równoległych z przesyłaniem komunikatów; może być zaimplementowany na maszynie szeregowej.
- Wykonuj wiele algorytmów ewolucyjnych równoległe (każdy z własną populacją na swojej wyspie). Wyspy połączone są siecią połączeń (zazwyczaj, pierścień albo torus).
- Po pewnej skończonej liczbie generacji algorytmu następuje **migracja**: pewna liczba osobników jest wysyłana z wyspy do sąsiednich wysp w sieci połączeń. Jednocześnie od każdego z sąsiadów jest odbierana ta sama liczba osobników.
- Nie ma gwarancji, że każda z populacji przeszukuje inny region przestrzeni rozwiązań.

Model wyspowy - parametry

- Jak często następuje migracja ?
 - Zbyt często → wszystkie populacje zbiegną do tego samego rozwiązania. Zbyt rzadko → tracimy czas.
 - Większość autorów stosuje migrację co 25-100 generacji.
- Jak dużo osobników wymieniać? Rekomendowana jest wymiana niewielkiej liczby (2-5) osobników.
- Które osobniki wysyłać (najlepsze, losowo wybrane).
- Czy wysyłany osobnik jest usuwany z populacji źródłowej ?
- W jaki sposób następuje inkorporacja osobników do populacji docelowej (na miejsce losowo wybranych, na miejsce najgorszych).
- Jaki przyjąć rozmiar podpopulacji (pewien minimalny rozmiar krytyczny jest niezbędny)

Model dyfuzyjny (ang. diffusion)



- sąsiad osobnika
- osobnik

- Na pojedynczą populację jest nałożona struktura przestrzenna (np graf typu grid). Jeden element populacji rezyduje w wierzchołku gridu.
- Dla każdego wierzchołka gridu zdefiniowane jest sąsiedztwo (np. zbiór wierzchołków odległych o 1 hop).
 - pojedynczy element umieszczony w wierzchołku jest zatem członkiem kilku nakładających się podpopulacji.
- Dla osobnika umieszczonego na wierzchołku gridu, selekcja, rekombinacja i sukcesja odbywa się w zbiorze sąsiadów.
 - Poszczególne podpopulacje przeszukują różne regiony przestrzeni rozwiązań, a dobre rozwiązania „dyfundują” przez grid.
- Technika zwana również **komórkowym** (ang. cellular) algorytmem ewolucyjnym. **7**

Model dyfuzyjny - implementacja

- W każdej generacji rozważamy wszystkie (nakładające się) sąsiedztwa. Dla każdego sąsiedztwa wykonujemy następujące operacje:
 - Wybierz dwóch rodziców z sąsiedztwa.
 - Stwórz jednego potomka przez krzyżowanie, poddaj go mutacji, oblicz jego dopasowanie.
 - Wybierz jednego osobnika z sąsiedztwa, i zastąp go potomkiem.
- Szczegóły implementacji
 - Wybór rodziców może być oparty na ich dopasowaniu albo losowy. Często jednym z rodziców jest obowiązkowo osobnik w węźle centralnym sąsiedztwa.
 - Powszechne jest zastępowanie przez potomka osobnika w węźle centralnym, ale występuje kilka wariantów, np. „zastąp zawsze” albo „zastąp gdy potomek jest lepszy”.
 - Schemat nadaje się do implementacji na maszynach równoległych, choć nie jest ona taka prosta jak w przypadku modelu wyspowego.

Współdzielenie dopasowania

- Dopasowanie osobnika ulega modyfikacji (przed selekcją) tak, aby alokować osobników do niszy proporcjonalnie do dopasowania niszy. Dopasowanie osobnika jest obniżane proporcjonalnie do liczby osobników w jego otoczeniu, co określa funkcja współdzielenia (ang. sharing function)
- Schemat stosowany z selekcją proporcjonalną.
- Rozważamy wszystkie pary osobników (x_i, x_j) , także (x_i, x_i) . Musimy mieć określoną miarę odległości $d(x_i, x_j)$ albo w przestrzeni genotypu (np. odległość Hamminga dla łańcuchów binarnych) albo w przestrzeni fenotypu (np. odległość Euklidesa pomiędzy dwoma wektorami)
- Modyfikujemy funkcję dopasowania:

$$f'(x_i) = \frac{f(x_i)}{\sum_{j=1}^S sh(d(x_i, x_j))}$$

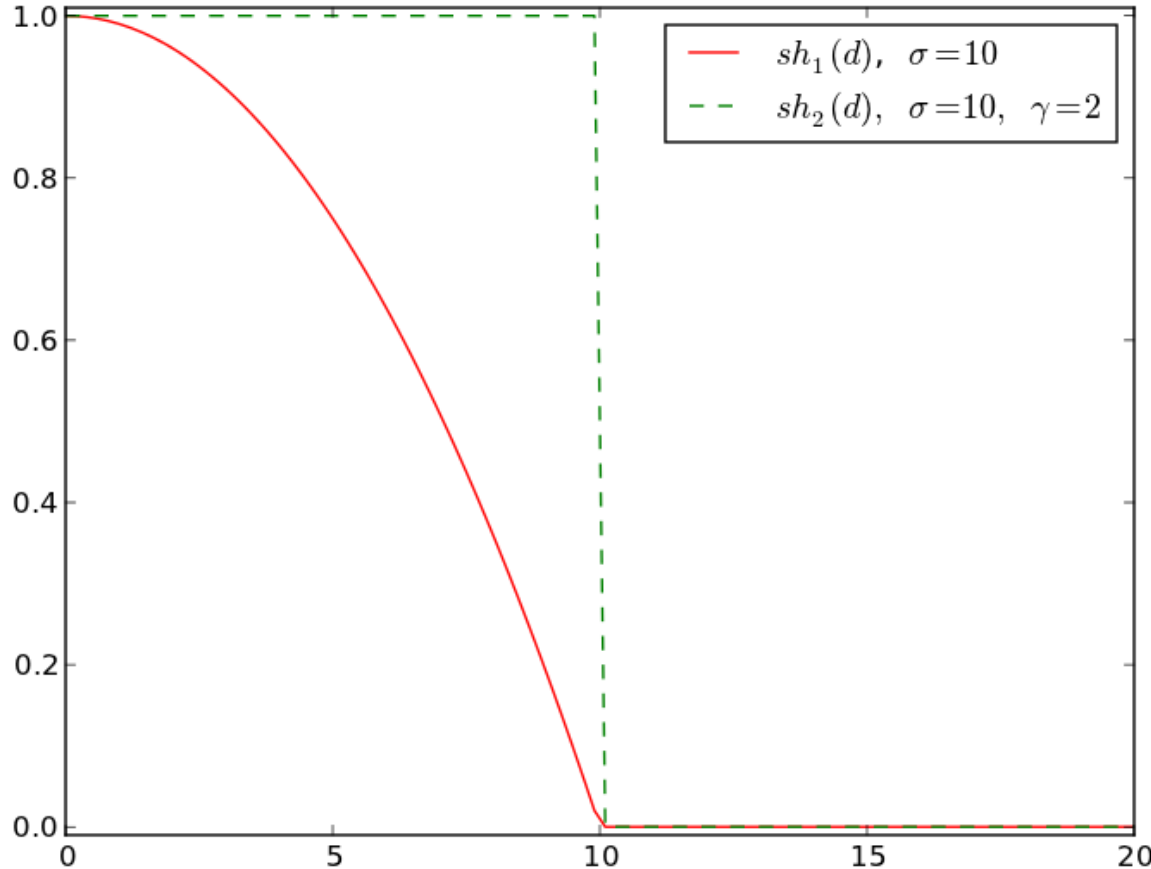
- gdzie S jest rozmiarem populacji, a sh jest **funkcją współdzielenia**, zależną od odległości dwóch osobników, która powinna spełniać warunki:

$$0 \leq sh(d) \leq 1$$

$$sh(0) = 1$$

$$\lim_{d \rightarrow \infty} sh(d) = 0$$

Funkcje współdzielenia



$$sh_2(d) = \begin{cases} 1 & \text{dla } d \leq \sigma \\ 0 & \text{dla } d > \sigma \end{cases}$$

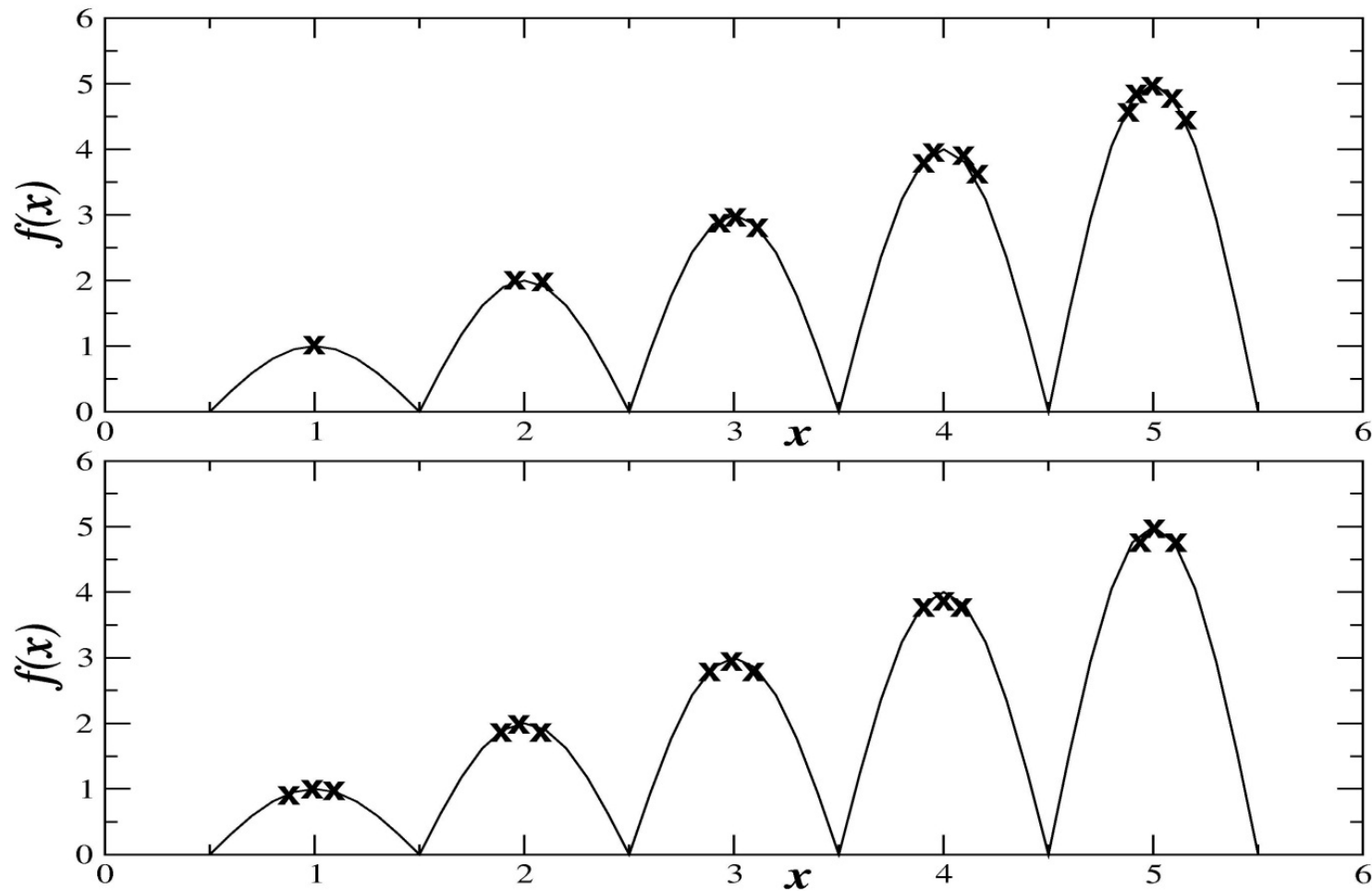
$$sh_1(d) = \begin{cases} 1 - (d/\sigma)^\gamma & \text{dla } d \leq \sigma \\ 0 & \text{dla } d > \sigma \end{cases}$$

- W przypadku funkcji sh_2 dopasowanie osobnika jest dzielone przez liczbę osobników w sąsiedztwie zdefiniowanym przez warunek $d \leq \sigma$.
- W przypadku funkcji sh_1 dalsi sąsiedzi mniej obniżają dopasowanie.
- Kluczowy parametr σ określa zasięg sąsiedztwa, osobniki poza sąsiedztwem nie obniżają dopasowania.

Ścisk (ang. crowding)

- Zaproponowane przez De Jonga (1975), zmodyfikowana wersja (deterministic crowding) opracowana przez Mahfouda (1992).
- Algorytm oparty na założeniu podobieństwa rodziców do potomków.
 - Podziel losowo populację rodziców na pary.
 - Wykonaj krzyżowanie w parach.
 - Wykonaj mutację, oblicz dopasowanie potomków.
 - Oznacz w każdej parze rodziców przez r_1, r_2 a potomków p_1, p_2 . Oblicz cztery odległości pomiędzy rodzicami i potomkami: $d(r_1, p_1)$, $d(r_1, p_2)$, $d(r_2, p_1)$, $d(r_2, p_2)$.
 - Zorganizuj turnieje pomiędzy rodzicami i potomkami tak aby suma odległości w turnieju była jak najmniejsza. To znaczy jeżeli $d(r_1, p_1) + d(r_2, p_2) < d(r_1, p_2) + d(r_2, p_1)$, to porównaj w turnieju r_1 z p_1 oraz r_2 z p_2 . W przeciwnym wypadku porównaj r_1 z p_2 oraz r_2 z p_1 .
 - Zwycięzcy turniejów przechodzą do nowej generacji.
- W przeciwieństwie do współdzielenia funkcji dopasowania, metoda nie wymaga parametrów.

Współdzielenie dopasowania i ścisk - porównanie (rys. Eiben i Smith, 2003)



- Ścisk rozdziela populację równomiernie na poszczególne nisze.
- W przypadku współdzielenia dopasowania, rozmiar podpopulacji jest proporcjonalny do dopasowania niszy.

Optymalizacja wielokryterialna

- W wielu praktycznych zastosowaniach, jakość rozwiązania jest definiowana w odniesieniu do kilku, najczęściej sprzecznych kryteriów.
- „Skonstruuj optymalny samochód”.
- W takim wypadku, zamiast przedstawiać użytkownikowi jedno „optymalne” rozwiązanie dobrze by było zaprezentować mu zbiór rozwiązań reprezentujący różne kompromisy pomiędzy kryteriami.
- Najprostsze rozwiązanie: **skalaryzacja**. Jeżeli masz dany zbiór n kryteriów f_1, f_2, \dots, f_n . Skonstruuj jedną funkcję dopasowania jako ważoną sumę n kryteriów:

$$f(x) = \sum_{i=1}^n \omega_i * f_i(x)$$

- zakłada że znamy preferencja użytkownika („spalanie jest 2 razy ważniejsze niż przyspieszenie”) zanim poznamy zakres możliwych rozwiązań.

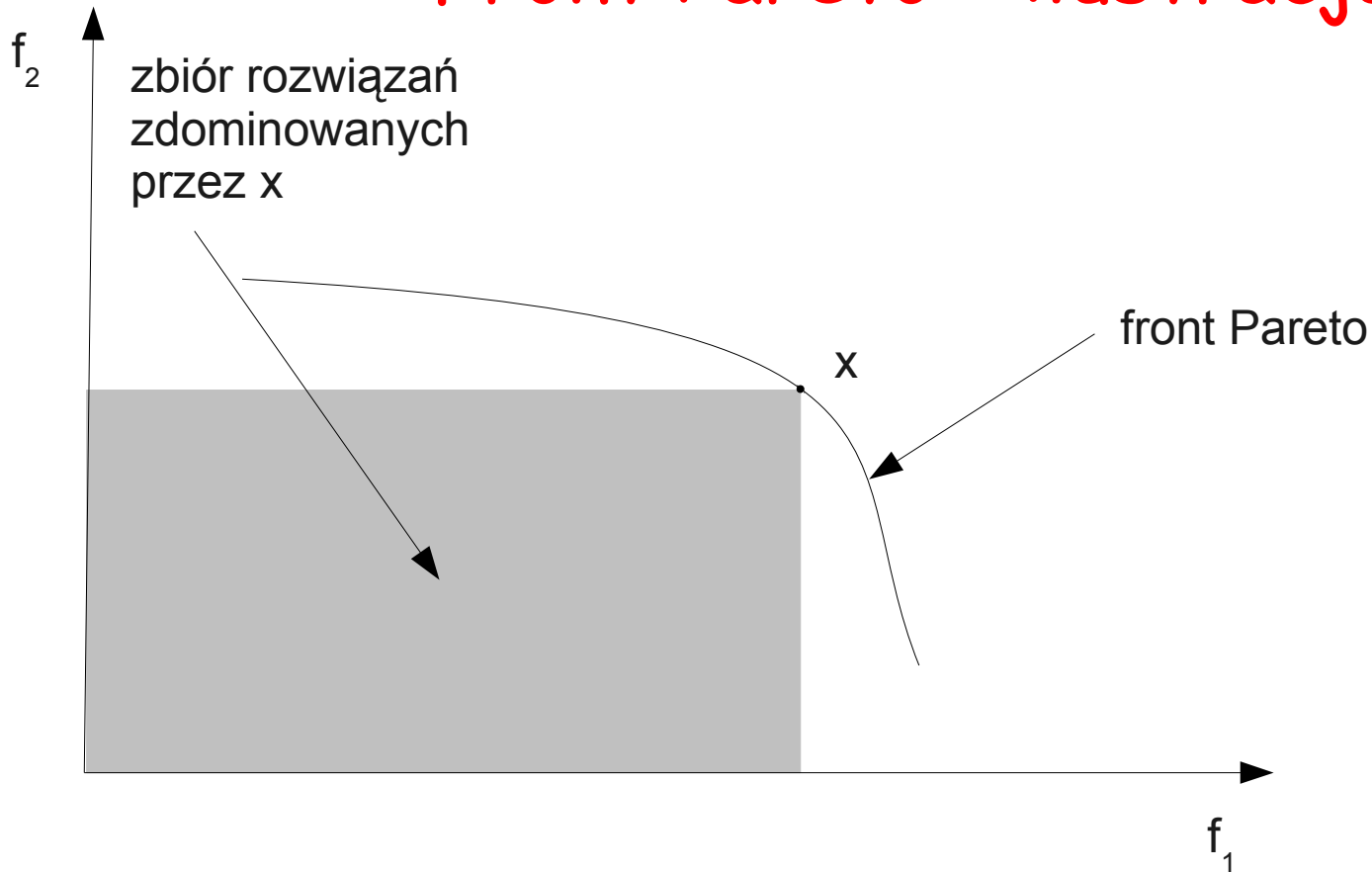
Pojęcie dominacji

- Dany jest zbiór funkcji n dopasowania (maksymalizacja) f_1, f_2, \dots, f_n . Rozwiązanie x dominuje nad rozwiązaniem y wtedy i tylko wtedy gdy:

$$\forall_{i \in \{1, 2, \dots, n\}} f_i(x) \geq f_i(y) \quad \text{oraz} \quad \exists_{i \in \{1, 2, \dots, n\}} f_i(x) > f_i(y)$$

- Innymi słowy x dominuje nad y , jeżeli jest przynajmniej tak samo dobre pod względem wszystkich kryteriów i lepsze pod względem przynajmniej jednego kryterium.
- Rozwiązanie x jest niezdominowane, jeżeli nie istnieje rozwiązanie y dominujące nad x .
- W przypadku nietrywialnego problemu optymalizacyjnego, w którym funkcje kryterialne konfliktują ze sobą, brak jest pojedynczego rozwiązania, które dominowałoby nad wszystkimi innymi.
- W takim przypadku interesuje nas znalezienie zbioru rozwiązań niezdominowanych, mających tę własność że nie można ich ulepszyć pod względem którejkolwiek z funkcji dopasowania nie pociągając ze sobą pogorszenia pozostałych.
- Ten zbiór rozwiązań niezdominowanych nazywany jest **frontem Pareto**.

Front Pareto - ilustracja

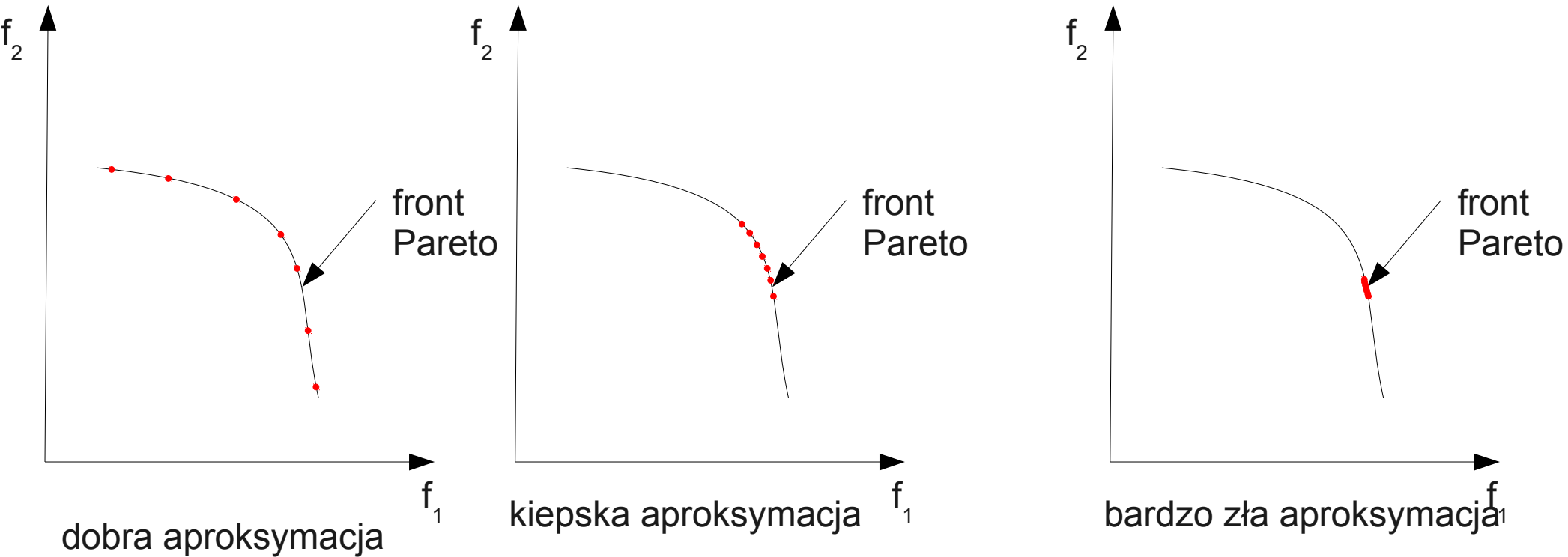


- Przetwarzanie populacji rozwiązań przez algorytm ewolucyjny sprawia, że mamy szansę jednoczesnego poszukiwania zbioru punktów aproksymujących front Pareto.

Reprodukcja oparta na dominacji

- Selekcja turniejowa: zwycięzcą turnieju jest osobnik, dla którego nie istnieje w turnieju osobnik go dominujący.
- Selekcja rangowa (wersja 1):
 - Znajdź w populacji osobniki niezdominowane, nadaj im rangę 0 (najwyższą) oraz pomiń je w dalszej części procedury.
 - Znajdź w populacji osobniki niezdominowane, nadaj im rangę 1, oraz pomiń je w dalszej części procedury.
 - Powtarzaj poprzedni krok, dopóki w populacji występują niepominięte osobniki.
- Selekcja rangowa (wersja 2):
 - Ranga osobnika x jest równa liczbie osobników w populacji, nad którymi x dominuje.

Aproksymacja frontu Pareto



- Potrzebne są mechanizmy utrzymywanie różnorodności w populacji.
 - Współdzielenie dopasowania (ale wymaga parametrów).
 - Rywalizacja potomka z najbliższym niezdominowanym osobnikiem w populacji (podobna do ścisku).
 - Użycie archiwum, zewnętrznego względem populacji, w którym przechowywane są niezdominowane punkty odkryte podczas przeszukiwania

Literatura dodatkowa

- E. Alba and M. Tomassini, “Parallelism and evolutionary algorithms,” IEEE Trans. Evol. Comput., vol. 6, no. 5, pp. 443–462, Oct. 2002. artykuł przeglądowy na temat zrównoleglania algorytmów ewolucyjnych.
- K. Deb, Multi-objective Optimization Using Evolutionary Algorithms, Wiley, 2001.
- C. A. Coello Coello, G. B. Lamont, D. A. van Veldhuizen, Evolutionary Algorithms for Solving Multi-Objective Problems, Kluwer, 2002.