

Tối ưu hóa (3)

TS. Đỗ Đức Đông
dongdoduc@gmail.com

Các cách tiếp cận

Tối ưu tổ hợp	Tối ưu liên tục
<ul style="list-style-type: none">• Các phương pháp truyền thống <p>Chứng minh hội tụ hoặc tỷ lệ tối ưu</p> <ul style="list-style-type: none">• Các phương pháp dựa trên thực nghiệm <p>+ Heuristic kiến trúc (constructive heuristic)</p> <p>+ Tìm kiếm địa phương (local search)</p> <p>+ Metaheuristics:</p> <ul style="list-style-type: none">- GA (Genetic Algorithms)- ACO (Ant Colony Optimization)- Memetic algorithm	<ul style="list-style-type: none">• Quy hoạch tuyến tính• Quy hoạch phi tuyến <p>+ Tìm kiếm địa phương</p> <ul style="list-style-type: none">- Phương pháp gradient <p>+ Tối ưu toàn cục</p> <ul style="list-style-type: none">- Quy hoạch lồi, hiệu lồi- Tìm kiếm ngẫu nhiên (Monter-Carlo)- GA (Genetic Algorithms) <p>...</p>

Thuật toán di truyền (Genetic Algorithm–GA)

- Đề xuất bởi John Holland, Đại học Michigan (năm 1975)
- Là phương pháp metaheuristic đang được sử dụng rộng rãi
- Thuật toán mô phỏng quá trình tiến hoá tự thích nghi của các quần thể sinh học dựa trên học thuyết Darwin để tìm lời giải các bài toán tối ưu
- Quá trình phát triển của mỗi quần thể tuân theo quy luật chọn lọc tự nhiên, tiến hoá qua các thế hệ kế tiếp nhau
- Các hậu duệ được sinh ra từ thế hệ trước thông qua quá trình sinh sản (di truyền và biến dị) cạnh tranh tự nhiên, cá thể nào thích nghi cao hơn với môi trường thì sẽ có khả năng lớn hơn trong tồn tại và sản sinh con cháu (cá thể nào “khỏe mạnh” sẽ sống sót và phát triển, các cá thể “yếu” sẽ bị loại bỏ).

Lược đồ thuật toán

{

$t \leftarrow 0;$

Khởi tạo quần thể $P(t)$;

while (chưa kết thúc)

{

$t \leftarrow t + 1;$

Chọn lọc $Q(t)$ từ $P(t)$; // nhờ bánh xe xổ số

Tái tạo $P(t)$ từ $Q(t)$;

Đánh giá $P(t)$ và chọn lọc cá thể tốt nhất;

}

}

Các bước cơ bản khi giải bài toán bằng thuật toán di truyền

- **Bước 1.** Chọn cấu trúc *nhiễm sắc thể* biểu diễn lời giải của bài toán.
- **Bước 2.** Khởi tạo tập lời giải ban đầu. Việc khởi tạo là ngẫu nhiên hay có thể áp dụng một vài thuật toán heuristic
- **Bước 3.** Chọn hàm đánh giá mức độ tốt của lời giải (để đánh giá các cá thể trong quần thể lời giải theo độ thích nghi của chúng).
- **Bước 4.** Thiết kế các toán tử di truyền. Các toán tử này được thiết kế nên dựa trên *nhiễm sắc thể*, đặc điểm bài toán và các ràng buộc của nó.
- **Bước 5:** Xác định các tham số cho bài toán. Các tham số này có thể không thay đổi trong quá trình tiến hoá hoặc có thể tự điều chỉnh tham số theo thời gian.

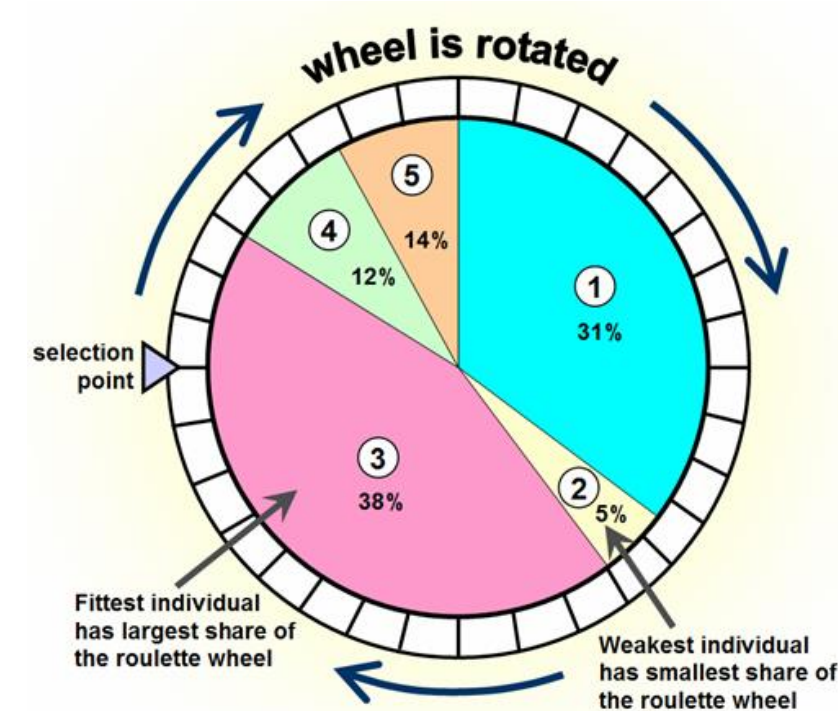
Chọn cấu trúc *nhiệm sắc thể*

Chọn cấu trúc *nhiệm sắc thể* biểu diễn lời giải của bài toán.

- Dãy bit (0101 ... 1100)
- Dãy số (43.2, -33.1, ... ,0.0, 89.2)
- Hoán vị (E11 E3 E7 ... E1 E15)
- Danh sách các luật (R1 R2 R3 ... R22 R23)
- Thành phần của chương trình (genetic programming)
- ... hoặc một cấu trúc dữ liệu nào đó...

Thủ tục chọn lọc

- Với mỗi quần thể $P(t - 1)$ gồm N nhiễm sắc thể : $P(t - 1) = \{v_1, \dots, v_N\}$ ta xây dựng *bánh xe xổ số* (roulette wheel) thực hiện quá trình chọn lọc:
 - Đánh giá độ phù hợp toàn phần: $F = \sum_{i=1}^N eval(v_i)$
 - Tính các xác suất chọn p_i của nhiễm sắc thể v_i : $p_i = \frac{eval(v_i)}{F}$
 - Tính xác suất tích lũy q_i của v_i : $q_i = \sum_{j=1}^i p_j$
- Quá trình chọn lọc quần thể $Q(t)$ từ $P(t - 1)$ dựa vào bánh xe xổ số được thực hiện theo cách sau:
 - Đối với mỗi số tự nhiên $k \in \{1, 2, \dots, N\}$ ta tạo một số ngẫu nhiên $r_k \in [0, 1]$.
 - Nếu $q_i \geq r_k > q_{i-1}$ thì chọn v_i thuộc $Q(t)$. Mỗi nhiễm sắc thể có thể được chọn nhiều lần và $Q(t)$ vẫn được xem là có N phần tử. Các cá thể có độ thích nghi lớn sẽ có khả năng được chọn nhiều hơn.



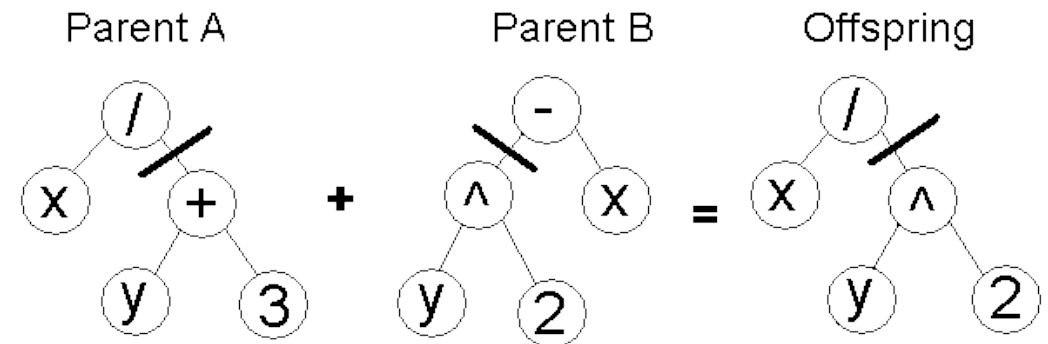
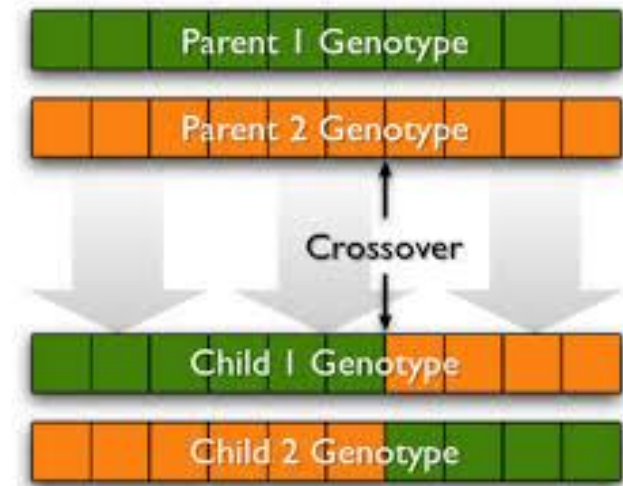
Toán tử di truyền (1)

Toán tử tương giao chéo (Crossover operator)

Với 2 nhiễm sắc thể $x = (x_1 \dots x_m)$ và $y = (y_1 \dots y_m)$, chọn điểm tương giao k (có thể ngẫu nhiên) ta sẽ sinh được hai nhiễm sắc thể mới:

➤ $x' = (x_1, \dots, x_k, y_{k+1} \dots y_m)$

➤ $y' = (y_1, \dots, y_k, x_{k+1} \dots x_m)$



Toán tử di truyền (2)

Toán tử biến dị (Mutation operator)

Với nhiễm sắc thể $x = (x_1, \dots, x_m)$ chọn điểm biến dị k (có thể ngẫu nhiên) ta sẽ sinh được nhiễm sắc thể mới.

Trước: (1 0 1 1 0 1 1 0)

Sau: (0 1 1 0 0 1 1 0)

Trước: (1.38 -69.4 326.44 0.1)

Sau: (1.38 -67.5 326.44 0.1)

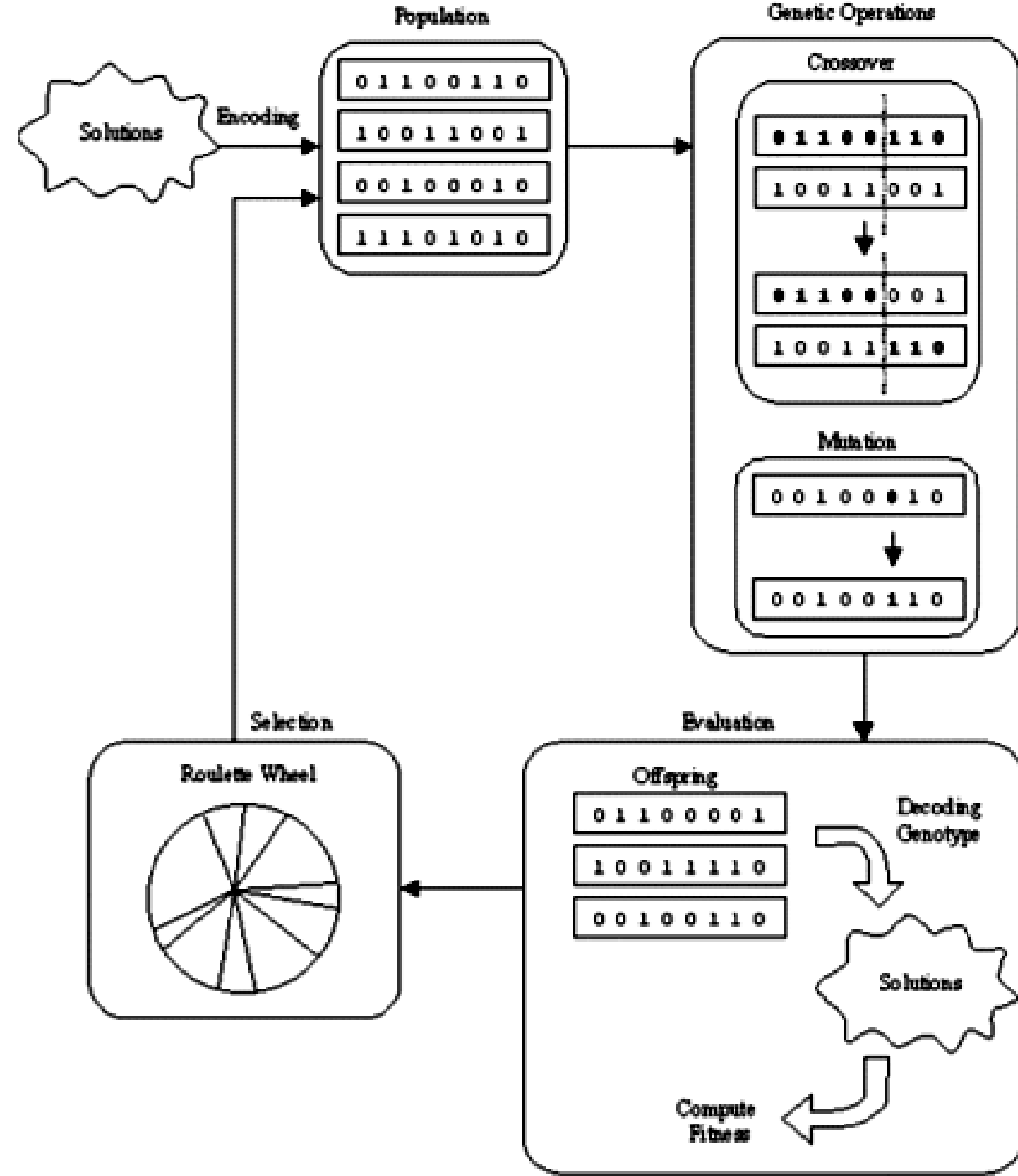
Thủ tục tái tạo

Cho trước các xác suất tương giao chéo p_c và xác suất biến dị p_m

- Đối với mỗi nhiễm sắc thể $v_i (i = 1, 2, \dots, n)$ thuộc $Q(t)$, tạo một số ngẫu nhiên $r \in [0, 1]$. Nếu $r < p_c$ thì v_i được đưa vào tập tương giao chéo.
- Sau khi tương giao chéo, đối với mỗi gene của nhiễm sắc thể ta tạo một số ngẫu nhiên $r \in [0, 1]$. Nếu $r < p_m$ thì gene này được biến dị.
- Quá trình trên cho ta quần thể $P(t)$ của thế hệ t và được đánh giá để chọn phần tử có giá trị thích nghi tốt nhất.

Ví dụ 1

- Cực đại hàm $f(x) = x^2$,
trong đó x nguyên thuộc $[0,255]$



Ví dụ 2

Bài toán người chào hàng (Traveling Salesman Problem)

Tìm hành trình

- Mỗi thành phố qua đúng một lần
- Tổng độ dài là nhỏ nhất

Cấu trúc nhiệm sắc thể

Mô tả bằng danh sách các thành phố (là một hoán vị)

1) London 3) Dunedin 5) Beijing 7) Tokyo
2) Venice 4) Singapore 6) Phoenix 8) Victoria

Danh sách 1 (3 5 7 2 1 6 4 8)

Danh sách 2 (2 5 7 6 8 1 3 4)

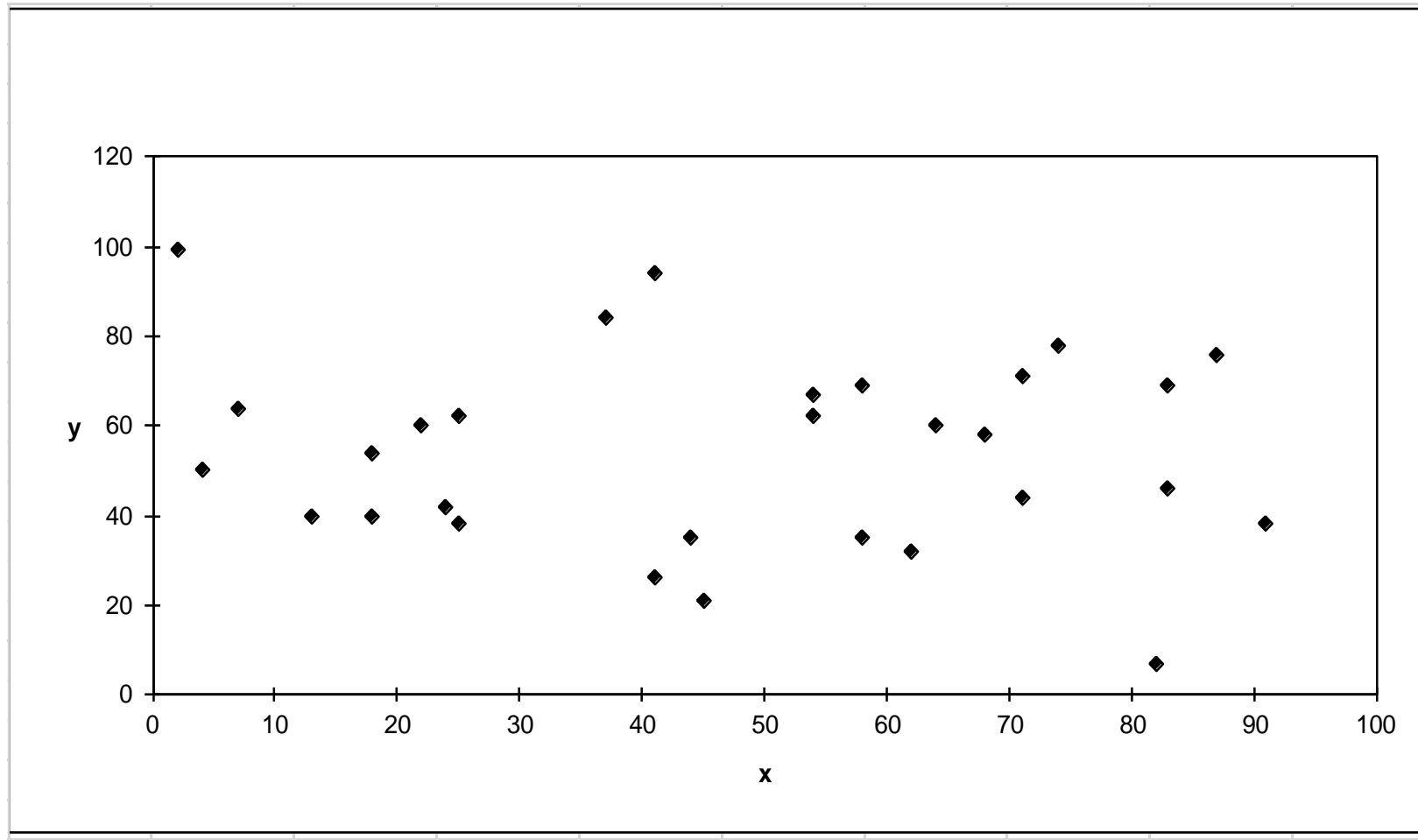
Toán tử tương giao chéo

			*		*		
Cha	(3	5	7	2	1	6	4 8)
Mẹ	(2	5	7	6	8	1	3 4)
<hr/>							
Con	(5	8	7	2	1	6	3 4)

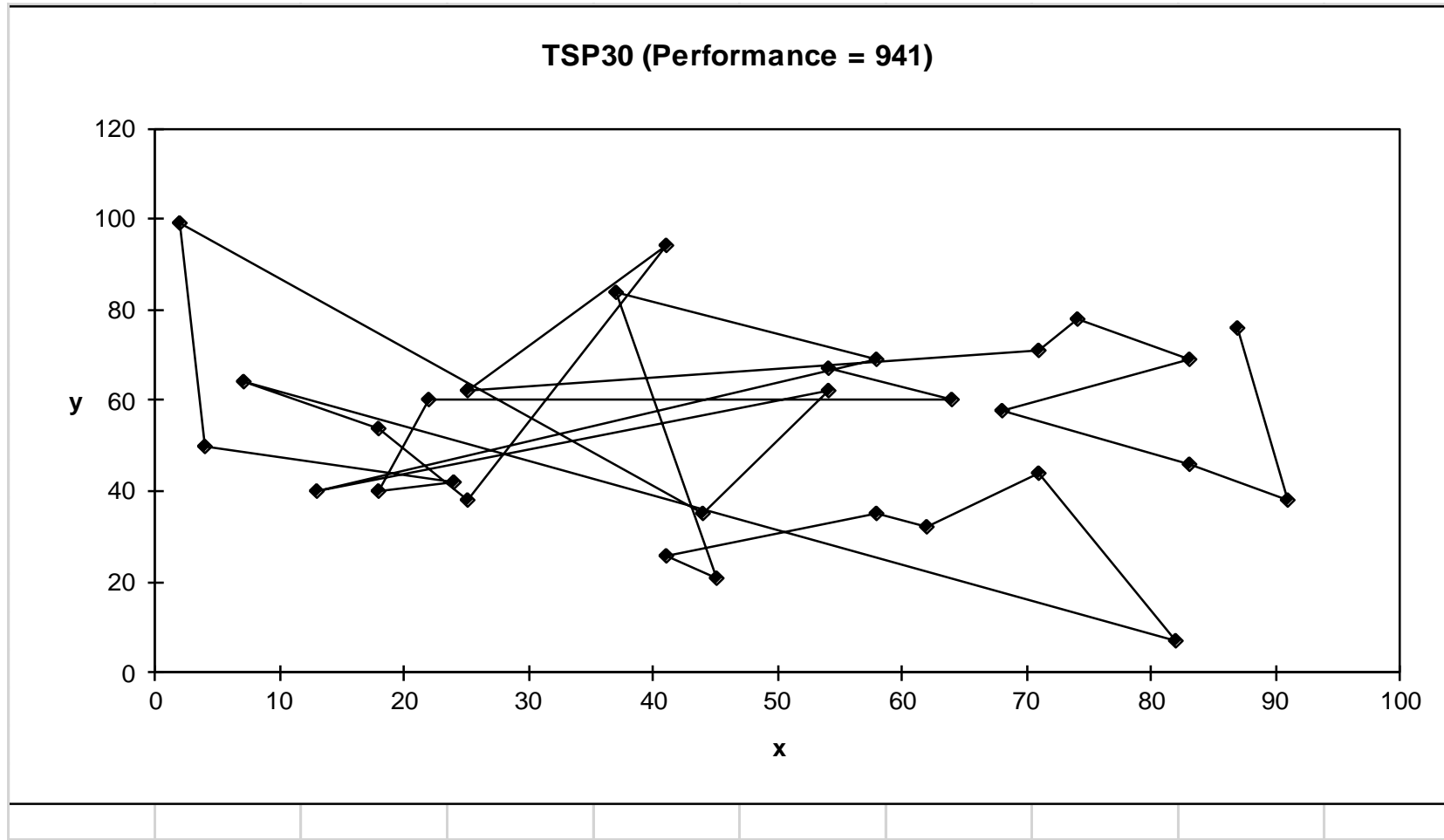
Đột biến

			*		*		
Trước:	(5	8	7	2	1	6	3 4)
Sau:	(5	8	6	2	1	7	3 4)

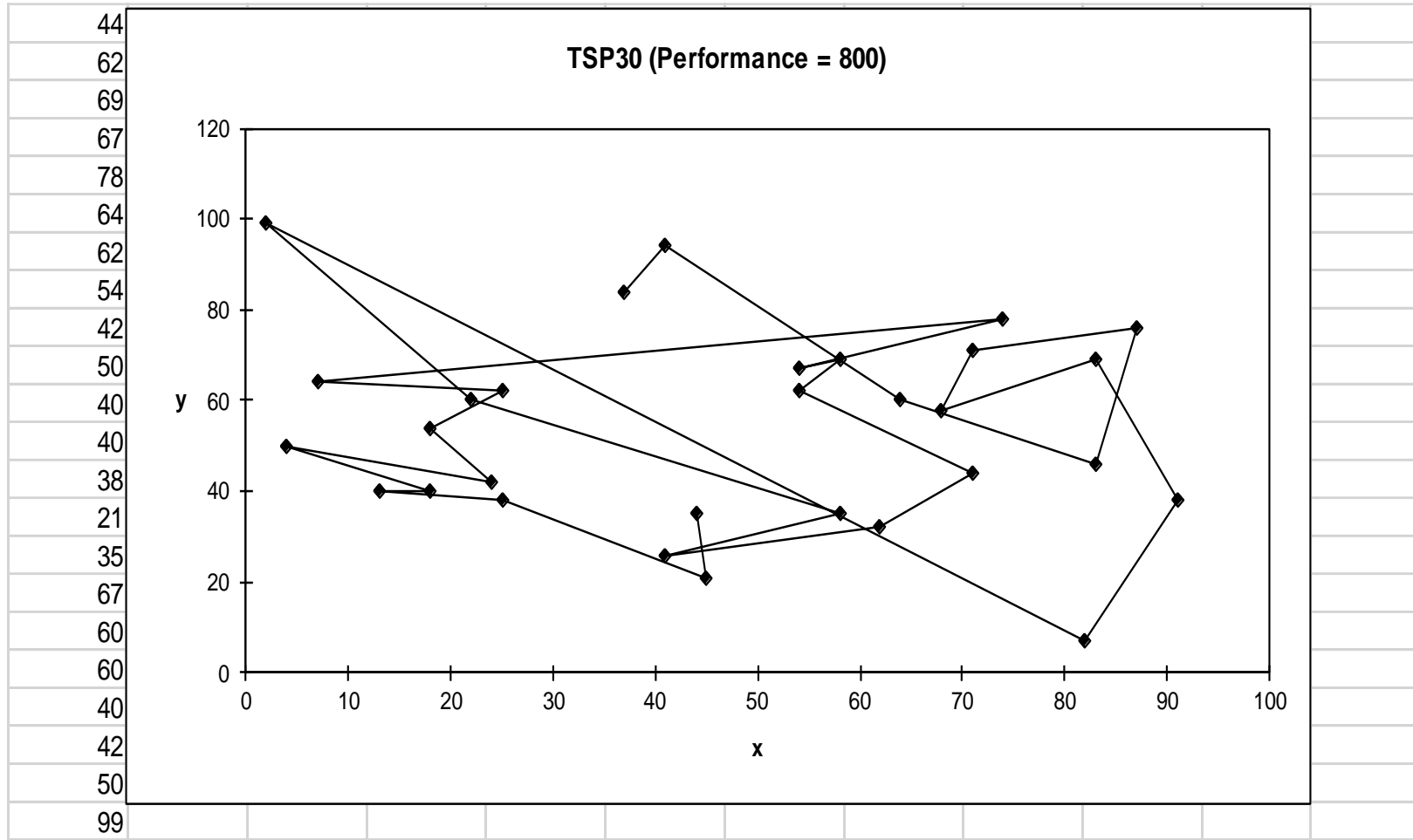
TSP Example: 30 Cities



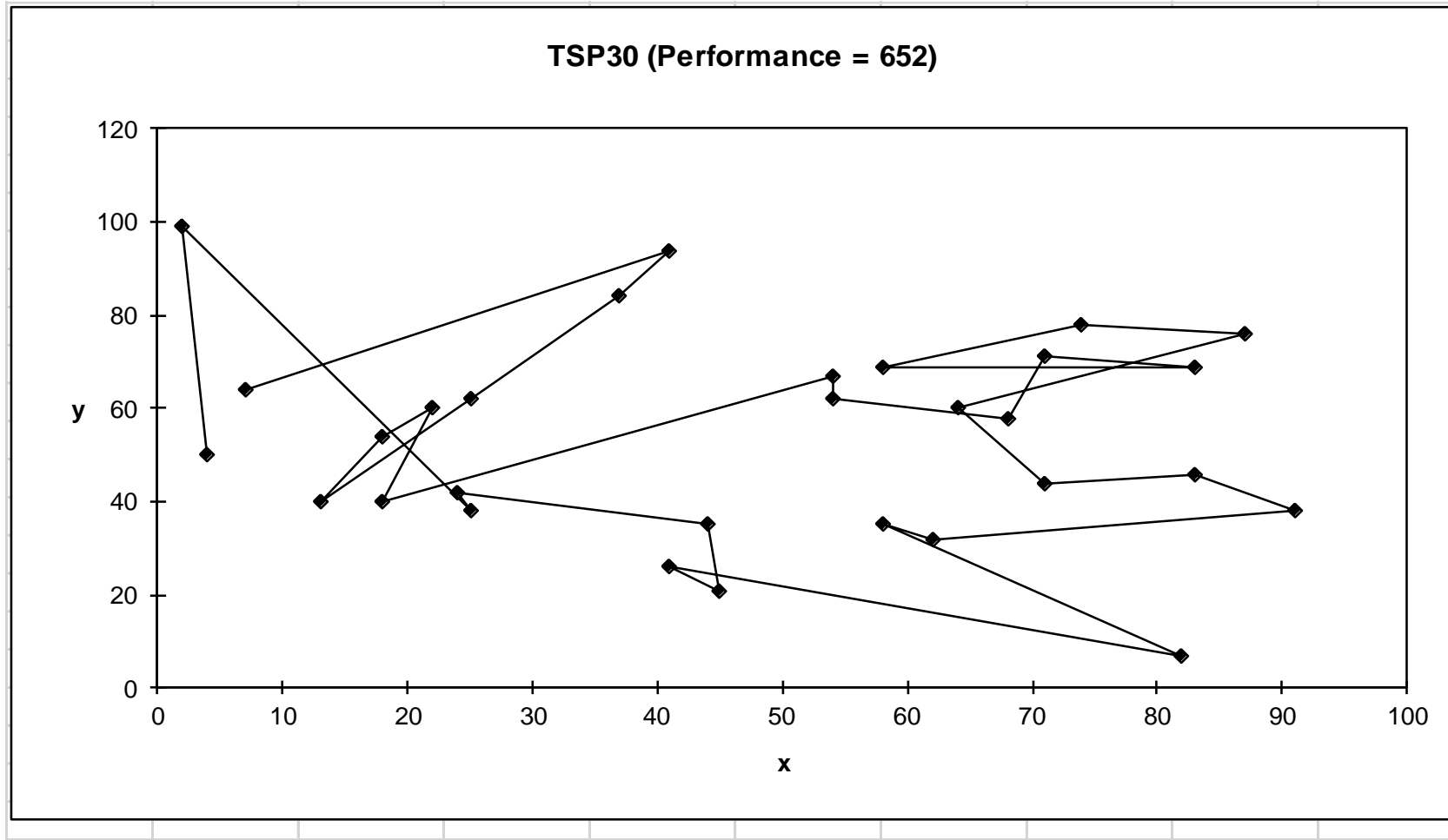
Solution (Distance = 941)



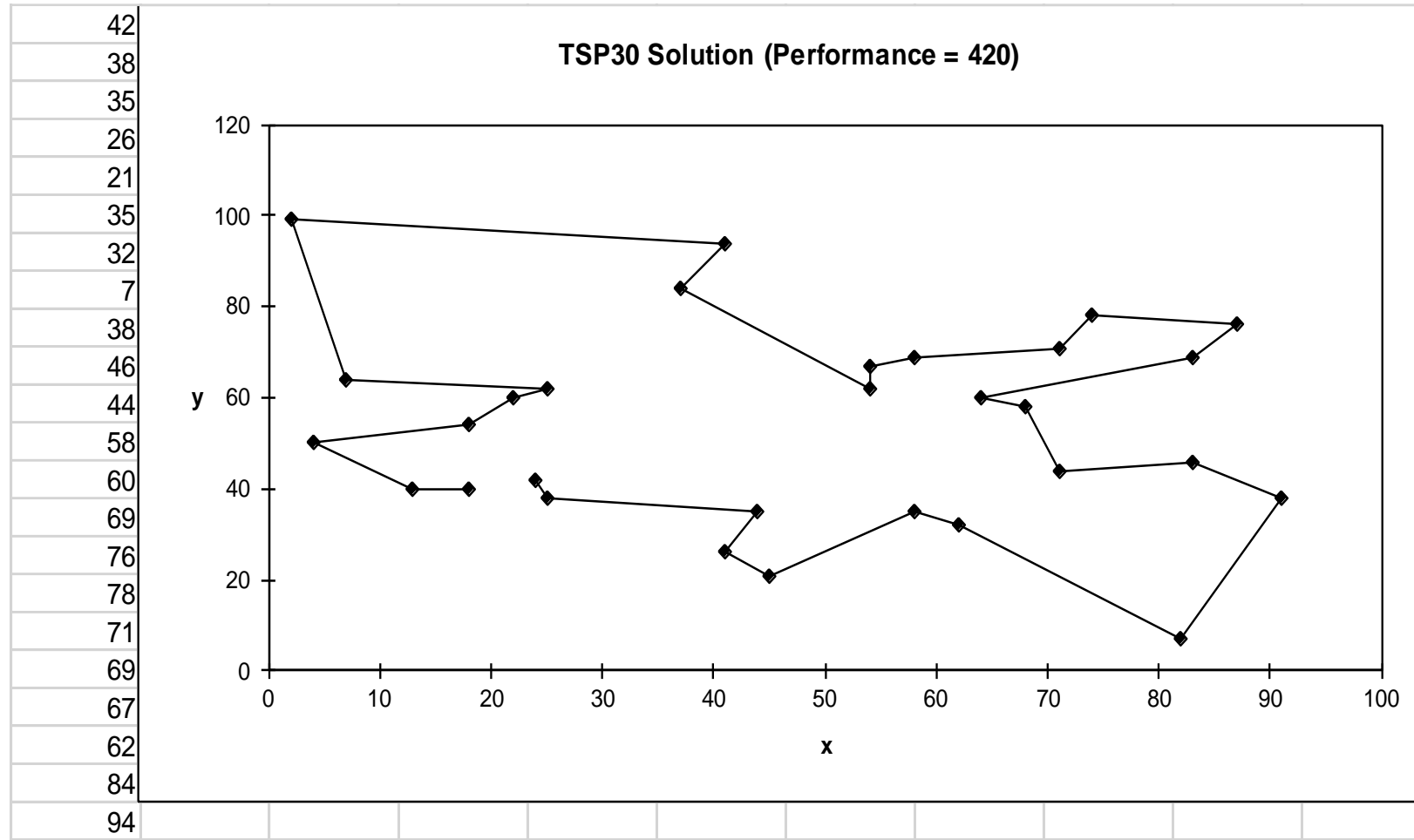
Solution (Distance = 800)



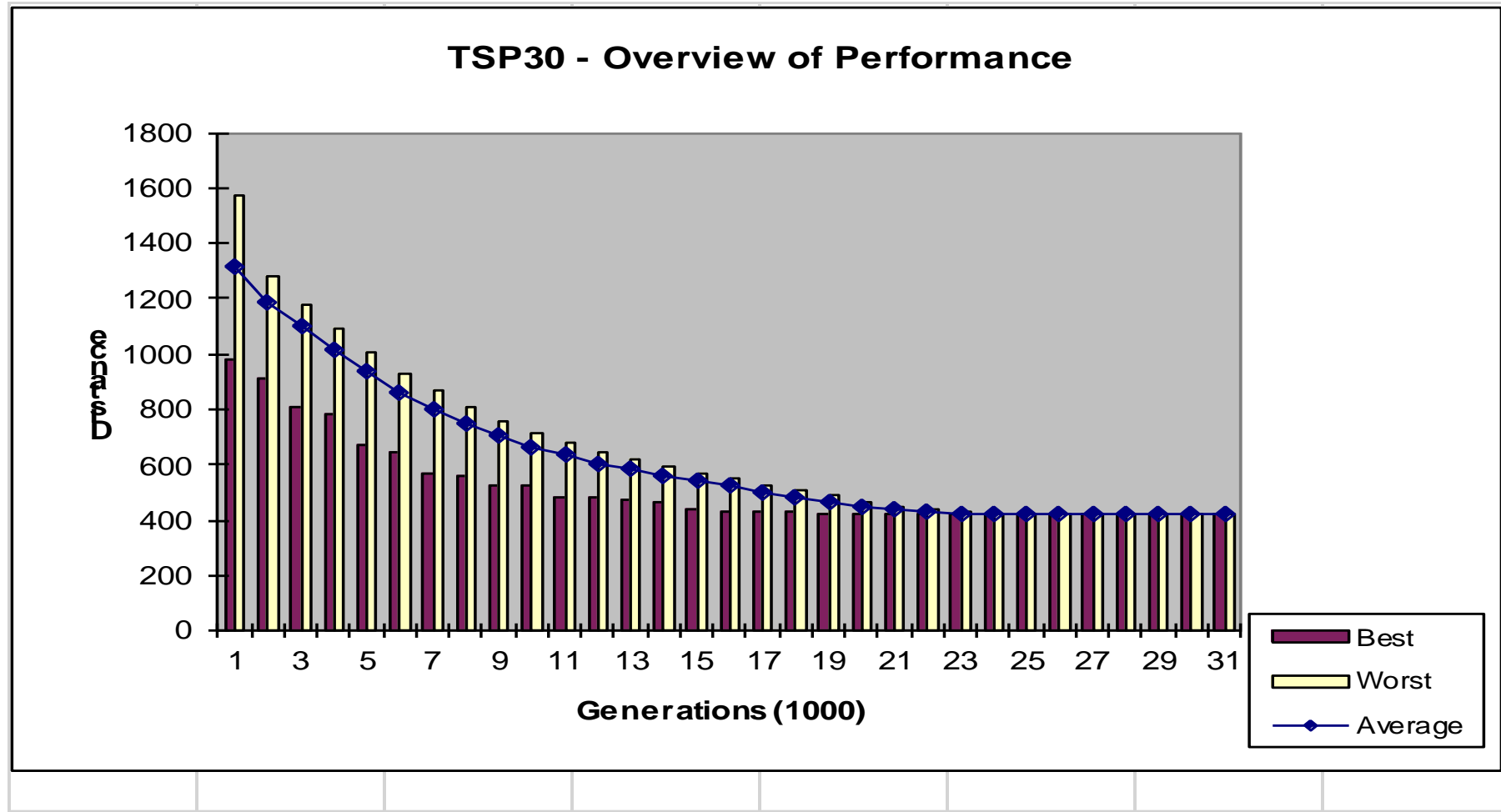
Solution (Distance = 652)



Best Solution (Distance = 420)

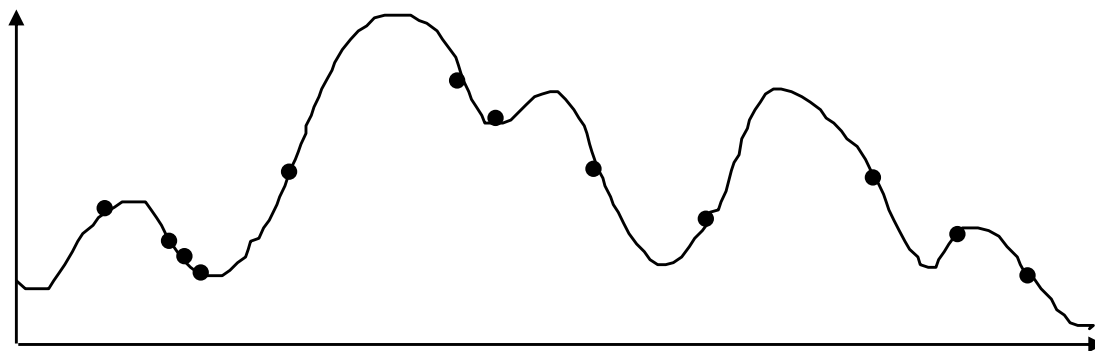


Overview of Performance

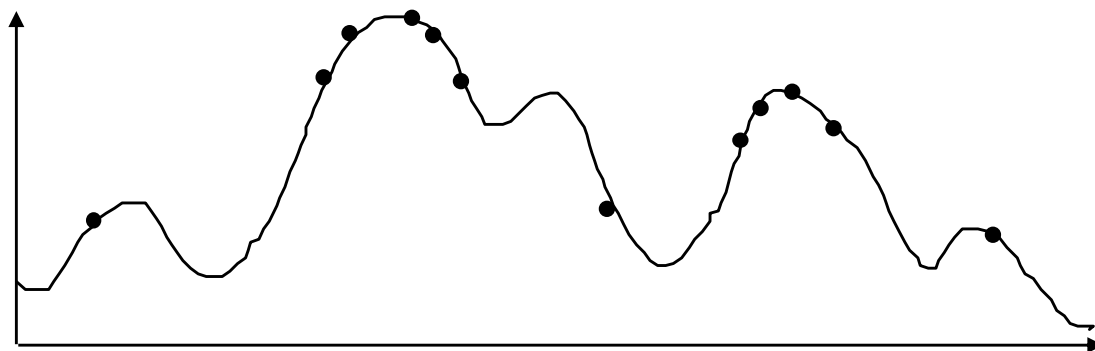


Nhận xét thuật toán di truyền

- Các kết quả nghiên cứu , đánh giá về sự hội tụ của GA còn rất nghèo, chỉ mới ở mức chứng minh sự hội tụ theo xác suất tới lời giải tối ưu của bài toán.
- Tuy nhiên, về mặt thực hành, giải thuật di truyền vẫn là một giải thuật được ưa thích để giải các bài toán khó trong thực tế và cho lời giải đủ tốt.
- Đặc biệt GA tỏ ra rất hiệu quả đối với các bài toán mà hàm mục tiêu phức tạp, có nhiều cực trị địa phương và không trơn. Đối với các bài toán đã có phương pháp giải tốt bằng phương pháp truyền thống thì GA vẫn kém hiệu quả hơn.



Phân bố của các cá thể trong thế hệ ban đầu



Phân bố của các cá thể trong thế hệ sau

Câu hỏi

- Điều kiện kết thúc của thuật toán?
- Hàm đánh giá độ thích nghi thỏa mãn điều kiện gì?
- Có thể sử dụng heuristic cấu trúc vào giai đoạn nào của thuật toán di truyền?
- Có thể sử dụng tìm kiếm cục bộ vào giai đoạn nào của thuật toán di truyền?
- Thuật toán di truyền gồm những tham số nào? Cách lựa chọn các tham số đó?