Jednowymiarowe zagadnienia minimalizacji.

Wcześniej zajmowaliśmy się przypadkiem, w którym zależność między wielkościami mierzonymi dało się przedstawić przy pomocy funkcji:

$$y(x) = a_1 + a_2 \cdot e^{-a_3 \cdot x}$$

Dopasowanie modelu do wyników pomiarów okazało się być problemem nieliniowym, prowadzącym do układu trzech równań

$$\begin{cases} a_1 \sum_{i=1}^{n} w_i + a_2 \sum_{i=1}^{n} w_i \cdot e^{-a_3 \cdot x_i} - \sum_{i=1}^{n} w_i \cdot y_i = 0 \\ a_1 \sum_{i=1}^{n} w_i \cdot e^{-a_3 \cdot x_i} + a_2 \sum_{i=1}^{n} w_i \cdot e^{-2 \cdot a_3 \cdot x_i} - \sum_{i=1}^{n} w_i \cdot y_i \cdot e^{-a_3 \cdot x_i} = 0 \\ a_1 \sum_{i=1}^{n} w_i \cdot x_i \cdot e^{-a_3 \cdot x_i} + a_2 \sum_{i=1}^{n} w_i \cdot x_i \cdot e^{-2 \cdot a_3 \cdot x_i} - \sum_{i=1}^{n} w_i \cdot x_i \cdot y_i \cdot e^{-a_3 \cdot x_i} = 0 \end{cases}$$

wyznaczających minimum funkcji χ^2

$$\chi^2(a_1, a_2, a_3) = \sum_{i=1}^{n} \left(\frac{y_i - a_1 - a_2 \cdot e^{-a_3 \cdot x_i}}{\sigma_i} \right)^2$$

Wprowadzimy dodatkowe oznaczenia:

$$w_i = \frac{1}{\sigma_i^2} \qquad S_w = \sum_{i=1}^{n} w_i$$

$$S_y = \sum_{i=1}^{n} w_i \cdot y_i \qquad S_{ax} = \sum_{i=1}^{n} w_i \cdot e^{-a_3 \cdot x_i}$$

$$S_{xax} = \sum_{i=1}^{n} w_i \cdot x_i \cdot e^{-a_3 \cdot x_i} \qquad S_{yax} = \sum_{i=1}^{n} w_i \cdot y_i \cdot e^{-a_3 \cdot x_i}$$

$$S_{xyax} = \sum_{i=1}^{n} w_i \cdot x_i \cdot y_i \cdot e^{-a_3 \cdot x_i} \qquad S_{axx} = \sum_{i=1}^{n} w_i \cdot e^{-2 \cdot a_3 \cdot x_i}$$

$$S_{xaxx} = \sum_{i=1}^{n} w_i \cdot x_i \cdot e^{-2 \cdot a_3 \cdot x_i}$$

Przy ich pomocy zapiszemy powyższy układ trzech równań w postaci:

$$\begin{cases} a_1 \cdot S_w + a_2 \cdot S_{ax} = S_y \\ a_1 \cdot S_{ax} + a_2 \cdot S_{axx} = S_{yax} \\ a_1 \cdot S_{xax} + a_2 \cdot S_{xaxx} - S_{xyax} = 0 \end{cases}$$

gdzie parametr a_3 jest ukryty w wyrazach postaci $S_{.a..}$ (w wykładnikach eksponent). Równania układu są nieliniowe tylko ze względu na ten jeden parametr a_3. Jeżeli przyjmiemy, że parametr ma ustalona wartość (np. stała rozpadu izotopu jest znana), to parametry a_1, a_2 możemy wyznaczyć rozwiązując, na przykład, dwa pierwsze równania:

$$a_1 = \frac{S_y S_{axx} - S_{yax} S_{ax}}{S_w S_{axx} - (S_{ax})^2}$$

$$a_2 = \frac{S_w S_{yax} - S_y S_{ax}}{S_w S_{axx} - (S_{ax})^2}$$

Po wstawieniu do trzeciego

$$\frac{S_y S_{axx} - S_{yax} S_{ax}}{S_w S_{axx} - (S_{ax})^2} \cdot S_{xax} + \frac{S_w S_{yax} - S_y S_{ax}}{S_w S_{axx} - (S_{ax})^2} \cdot S_{xaxx} - S_{xyax} = 0$$

i uproszczeniu otrzymujemy

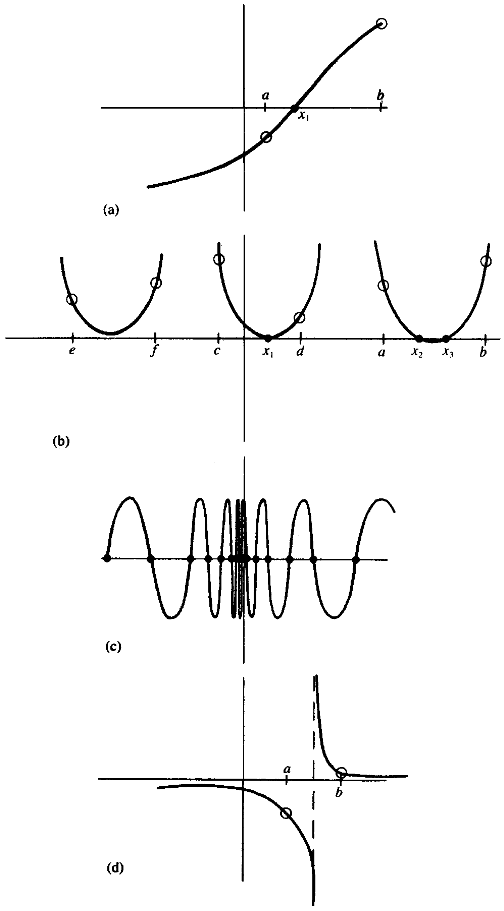
$$\left(S_y S_{axx} - S_{yax} S_{ax}\right) \cdot S_{xax} + \left(S_w S_{yax} - S_y S_{ax}\right) \cdot S_{xaxx} - \left(S_w S_{axx} - S_{ax}^2\right) \cdot S_{xyax} = 0$$

W naszym przypadku wyniki pomiarów i ich niepewności są ustalone, czyli lewa strona równania jest nieliniową funkcją jednej zmiennej

$$f\left(a_3; \{x_i, y_i, \sigma_i\}\right) = 0$$

Zatem trójwymiarowe zagadnienie udało się sprowadzić do problemu wyznaczenia miejsc zerowych funkcji jednej zmiennej.

Miejsca zerowe funkcji jednej zmiennej



(a) – pojedynczy pierwiastek ograniczony w przedziale (a,b)

(b) – w przypadku podwójnego pierwiastka funkcja może mieć ten sam znak na końcach przedziału

(c) – przykład funkcji z wieloma miejscami zerowymi

(d) – przykład funkcji zmieniającej znak w przedziale (a,b), ale nie mającej w nim miejsca zerowego.

Metoda bisekcji

Przed jej zastosowaniem, należy znaleźć przedział (a,b), który ogranicza miejsce zerowe, to znaczy taki, że $f(a)\cdot f(b)<0$.

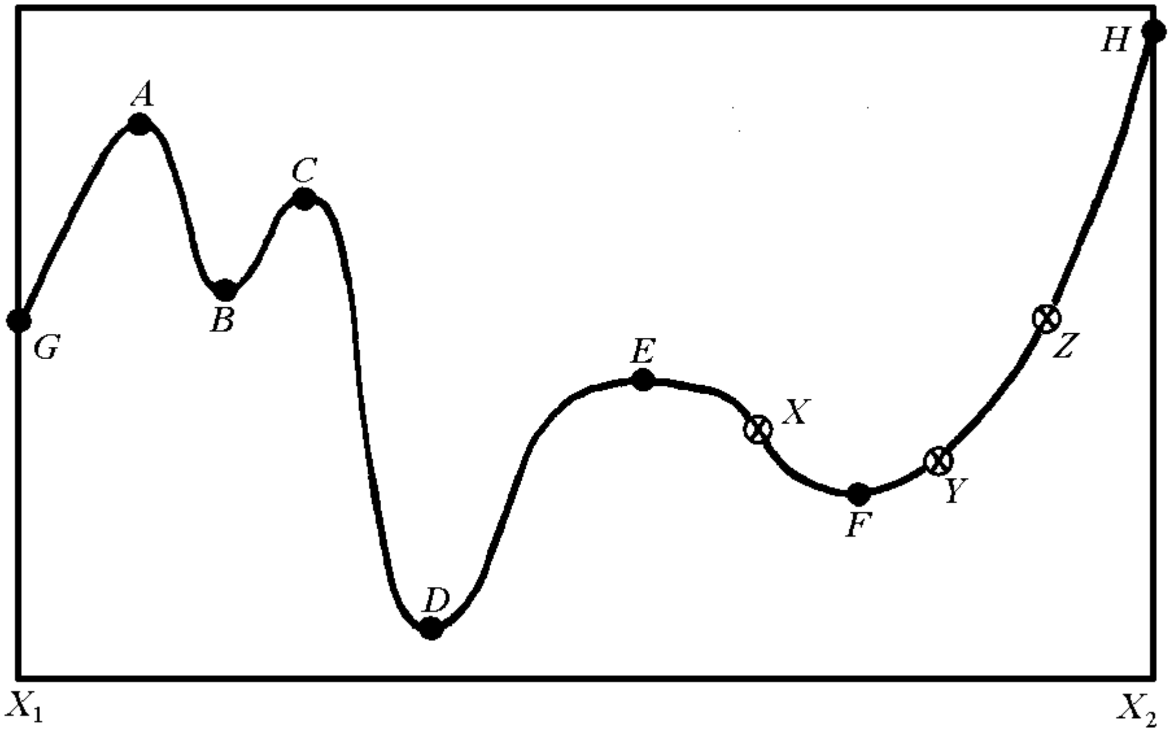
1. dzielimy przedział na pół $c=\dfrac{a+b}{2}$ i obliczamy $f(c)$
2. jeżeli $f(a)\cdot f(c)<0$, to podstawiamy $b=c$ (w przeciwnym przypadku $a=c$) i wracamy do 1.
3. kończymy jeżeli $f(c)=0$ lub $|a-b|$ jest dostatecznie małe.

Metoda jest pewna, przynajmniej jedno miejsce zerowe znajduje się w końcowym przedziale i stosunkowo szybko zbieżna. Po n krokach szerokość przedziału zawierającego miejsce zerowe wynosi

$$d_n=\frac{d_0}{2^n}$$

Na końcu musimy się jeszcze upewnić, czy nie znaleźliśmy przypadkiem lokalnego maksimum funkcji χ^2.

Bezpośrednia minimalizacja funkcji χ^2



A, C, E – lokalne maksima funkcji; H – globalne maksimum w przedziale (X_1, X_2).
G, B, F – lokalne minima w przedziale; D – globalne minimum funkcji w przedziale (X_1, X_2).

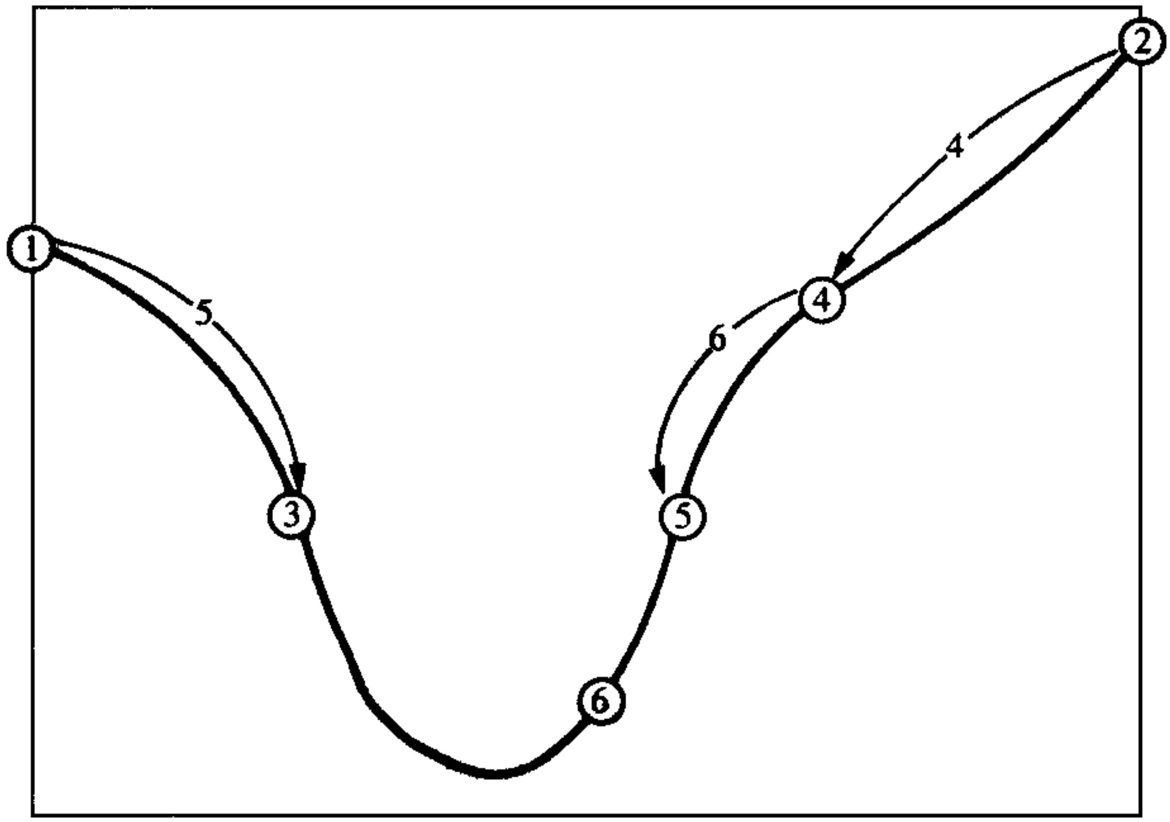
Metoda złotego podziału

Tryplet punktów $X < Y < Z$ ogranicza minimum funkcji jeżeli

$$f(Y) < f(X) \text{ i } f(Y) < f(Z)$$

Odpowiednikiem metody bisekcji jest następujący algorytm lokalizacji minimum funkcji.

1. Mamy tryplet $a < b < c$ ograniczający minimum funkcji $f(x)$
2. Obliczamy wartość $f(x)$ w punkcie $b < x < c$
3. Jeżeli $f(b) < f(x)$, to nowym trypletem jest $a < b < x$, w przeciwnym przypadku tym trypletem jest $b < x < c$
4. Znaleziony tryplet ograniczający minimum funkcji zastępuje początkowy w punkcie 1. i cykl powtarzamy
5. Kończymy jeżeli szerokość trypletu $c - a$ jest odpowiednio mała.



Kolejne kroki lokalizacji minimum. Początkowo minimum jest ograniczone przez (1, 3, 2). Funkcja jest obliczona w 4, który zastępuje 2 (1, 3, 4); następnie w 5, który zastępuje 1 (3, 5, 4); następnie w 6, który zastępuje 4. Teraz punkty (3, 6, 5) są trypletem zawierającym minimum.

Optymalizacja podziału

Środkowy punkt trypletu dzieli go tak, że b wyznacza ułamek W drogi od a do c

$$\frac{b-a}{c-a}=W \quad \frac{c-b}{c-a}=1-W\,.$$

Kolejny punkt x leży o dodatkowy ułamek Z za b

$$\frac{x-b}{c-a}=Z$$

Następny tryplet będzie miał szerokość $W+Z$ albo $1-W$ w stosunku do poprzedniego. Jeżeli chcemy zminimalizować szerokość nowego trypletu w najgorszym przypadku, to obie szerokości powinny być takie same

$$W+Z=1-W$$

albo

$$Z=1-2W$$

Oznacza to, że x powinien leżeć w odcinku (a, c) symetrycznie do b i wypadać w większej części tego odcinka. Jeżeli poprzedni podział też był optymalny ze względu na najgorszy przypadek, to x dzieli odcinek (b, c) w takim stosunku jak b dzieli (a, c)

$$\frac{Z}{1-W}=W$$

Wartość W wyznacza równanie kwadratowe

$$W^2-3W+1=0$$

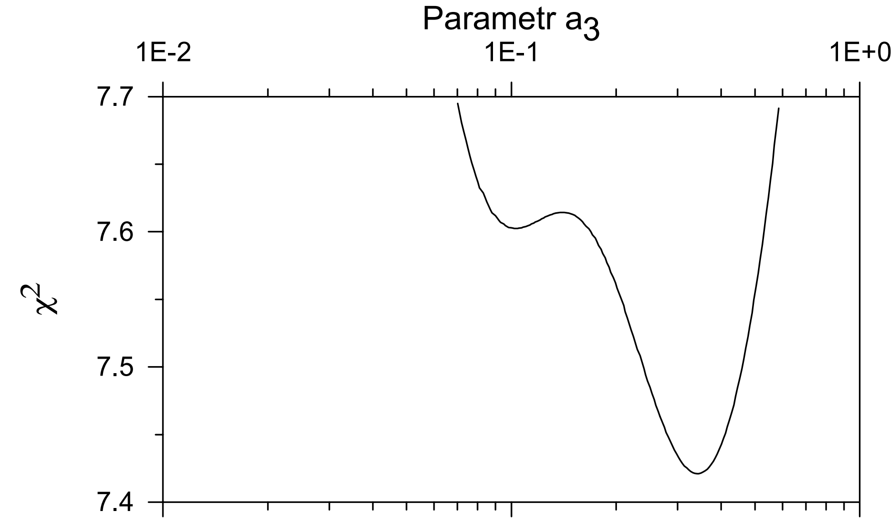
co daje

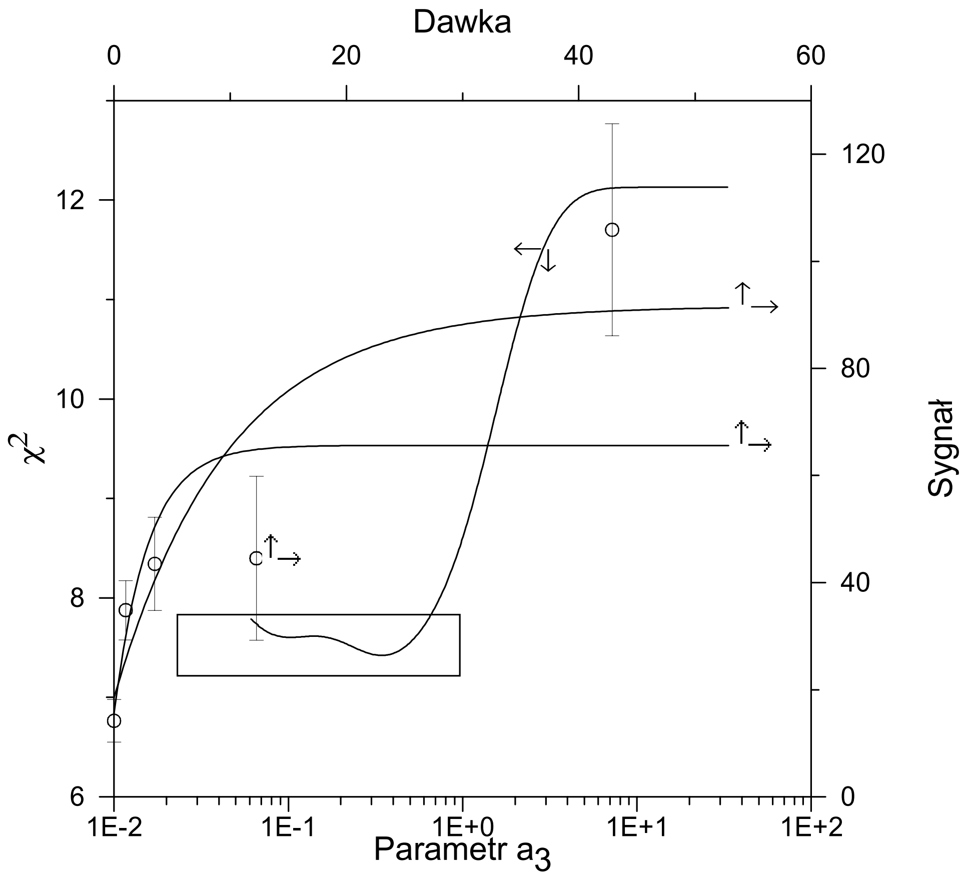
$$W=\frac{3-\sqrt{5}}{2}\approx 0{,}38197\,.$$

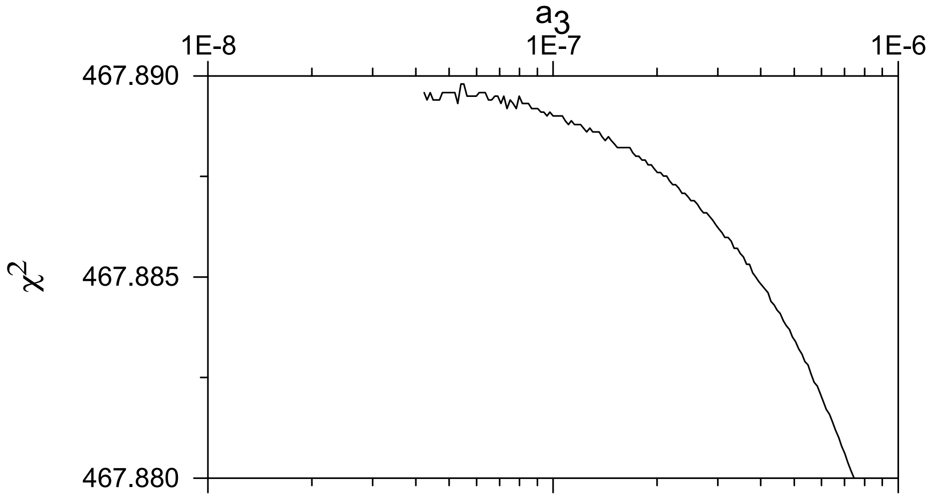
Optymalny tryplet ma zatem punkt środkowy we względnej odległości 0,38197 od jednego końca i 0,61803 od drugiego. Taki punkt nazywa się *złotym środkiem*, a podział *złotym podziałem*.

Przykład minimalizacji jednowymiarowej

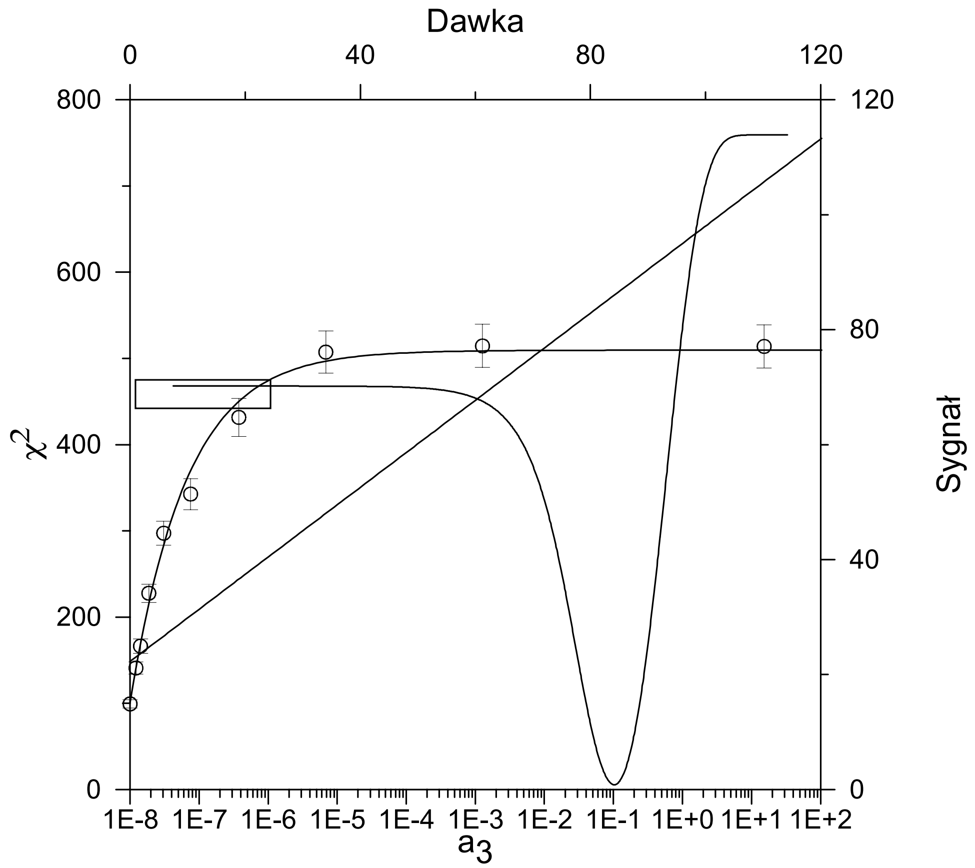
$$y(x) = a_1 + a_2 \cdot e^{-a_3 \cdot x}$$







a3
1E-8
1E-7
1E-6
467.890
467.885
467.880
χ2



Dawka
0
40
80
120
800
600
400
200
0
χ2
120
80
40
0
Sygnał
1E-8 1E-7 1E-6 1E-5 1E-4 1E-3 1E-2 1E-1 1E+0 1E+1 1E+2
a3

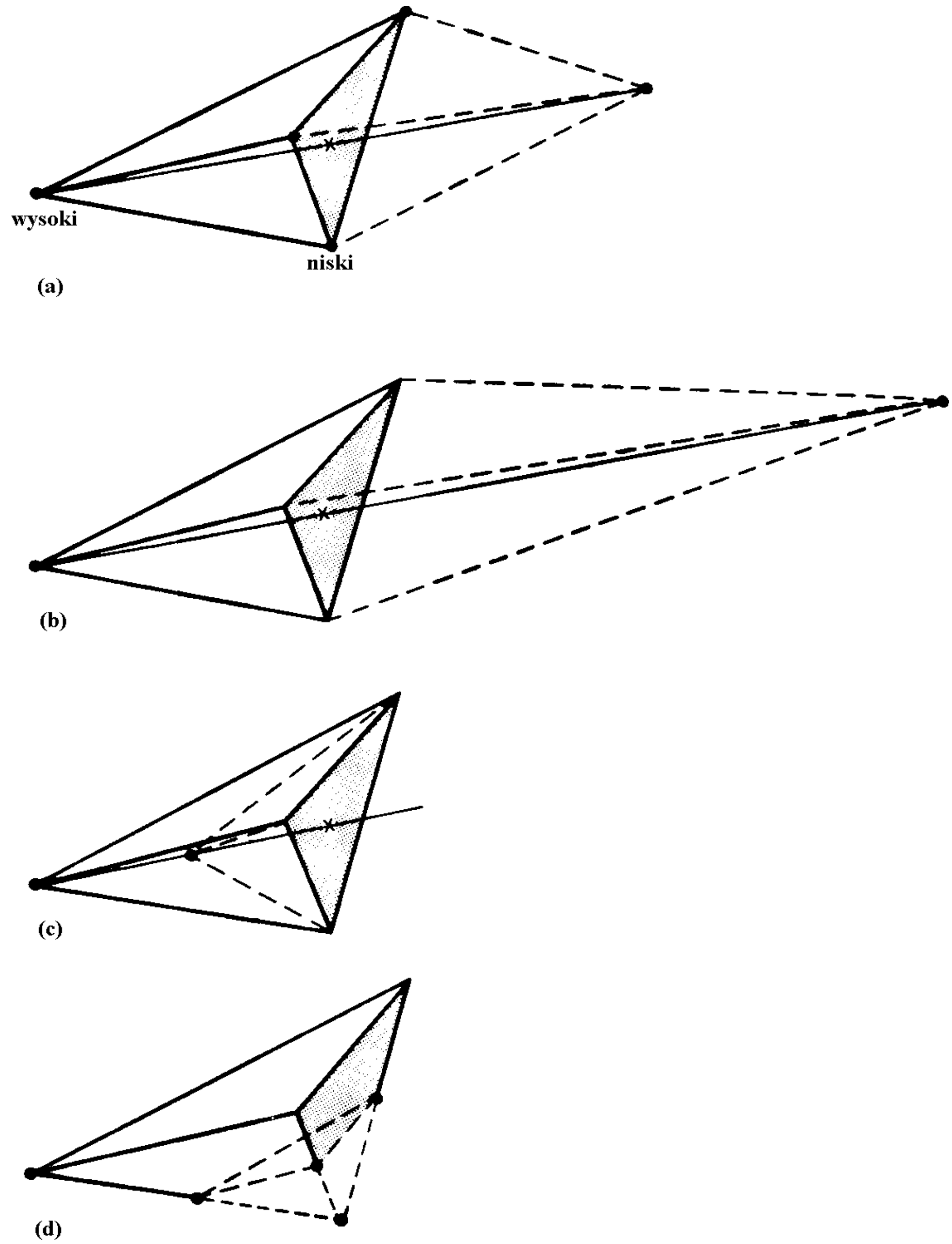
Metoda simpleks minimalizacji w wielu wymiarach

Simpleks jest tworem geometrycznym, który w przestrzeni n-wymiarowej składa się z $n + 1$ punktów (wierzchołków) oraz wszystkich łączących je odcinków, wielokątów, itd. Na płaszczyźnie simpleks jest trójkątem, w przestrzeni 3-wymiarowej czworościanem. Simpleks niezdegenerowany, to taki, który zawiera skończoną (n-wymiarową) objętość. Jeżeli jeden z wierzchołków niezdegenerowanego simpleksu obierzemy jako początek układu, to pozostałe wierzchołki wyznaczają n liniowo niezależnych wektorów w n-wymiarowej przestrzeni.

Metoda minimalizacji wymaga wybrania początkowego simpleksu o $n + 1$ wierzchołkach. Możemy, np. wybrać jeden punkt w przestrzeni parametrów $\mathbf{a}_0$, a pozostałe utworzyć tak

$$\mathbf{a}_i = \mathbf{a}_0 + \lambda \mathbf{e}_i$$

W kolejnych krokach simpleks jest przemieszczany w kierunku minimum (‘w dół zbocza’). W większości przypadków polega to na tym, że najgorszy wierzchołek (z największą wartością funkcji) jest przesuwany w kierunku przeciwległej ścianki do miejsca, gdzie wartość funkcji jest mniejsza. Jeżeli to nie daje efektu, to simpleks zostaje zmniejszany przez ściągnięcie wierzchołków w kierunku najlepszego wierzchołka (z najmniejszą wartością funkcji).



Możliwe wyniki kolejnego kroku simpleksu. Poprzedni simpleks (czworościan) jest narysowany ciągłą linią. ‘Wysoki’ i ‘niski’ są najgorszym i najlepszym wierzchołkiem simpleksu. Simpleks następny jest narysowany linia przerywaną.

(a) – odbicie najgorszego punktu względem przeciwległej ścianki
(b) – jak (a) i rozciągnięcie simpleksu w tym kierunku
(c) – zmniejszenie simpleksu w jednym wymiarze przez przesunięcie wysokiego wierzchołka
(d) – zmniejszenie simpleksu we wszystkich wymiarach przez ściągnięcie wierzchołków w kierunku najlepszego wierzchołka.