

Dopasowanie dowolnej funkcji do danych pomiarowych.

Do tej pory staraliśmy się dopasować do wyników pomiarów funkcje o ogólnej postaci:

$$y(x) = \sum_{k=1}^m a_k f_k(x),$$

zawierające m nieznanymi parametrów $a_1 \dots a_m$. Zakładaliśmy przy tym, że same funkcje $f_k(x)$ nie zawierają tych parametrów. Minimalizacja funkcji χ^2 (MNK) prowadziła do układu m równań liniowych ze względu na te parametry, które można rozwiązać metodami algebraicznymi. Dlatego zagadnienie dopasowania funkcji $y(x) = \sum a_k f_k(x)$ do wyników pomiarów nazywa się zagadnieniem liniowym.

W przypadku dopasowania dowolnej funkcji układ równań określających warunek minimum χ^2 jest najczęściej nieliniowy względem szukanych parametrów a_1, \dots, a_m . Taki układ nie może być rozwiązany metodami ścisłymi.

Metoda najmniejszych kwadratów (MNK)

Funkcja wiarygodności w takim przypadku ma ogólną postać

$$P(a_1, \dots, a_m; \{x_i, y_i, \sigma_{y_i}\}) = \prod_{i=1}^n \left(\frac{1}{\sigma_{y_i} \sqrt{2\pi}} \right) \cdot \exp \left[-\frac{1}{2} \sum_{i=1}^n \left(\frac{y_i - y(x_i)}{\sigma_{y_i}} \right)^2 \right]$$

gdzie $y(x)$ jest dowolną funkcją mającą m parametrów a_1, a_2, \dots, a_m , które dopasowujemy tak, żeby zmaksymalizować prawdopodobieństwo P .

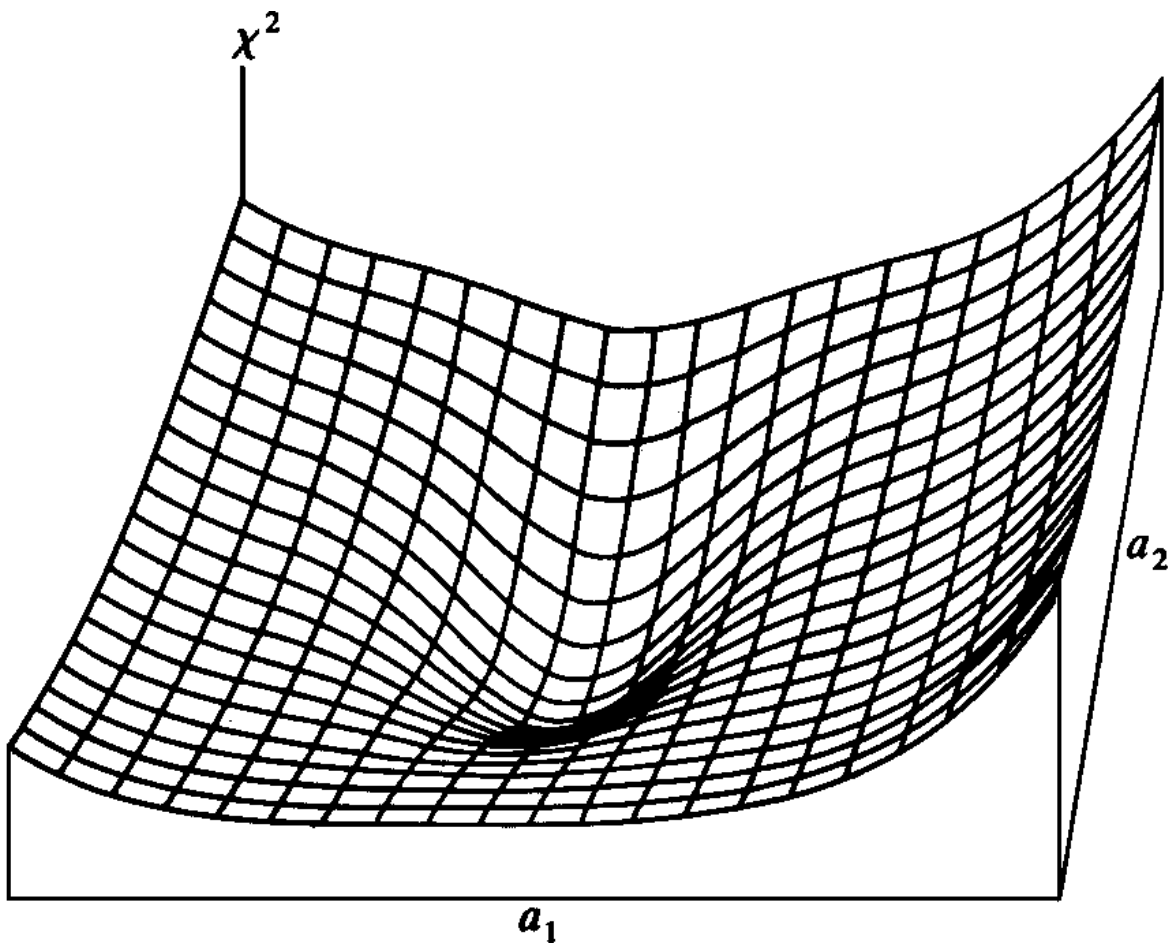
Maksymalizacji prawdopodobieństwa P odpowiada minimalizacja wykładnika eksponenty, czyli χ^2 .

$$\chi^2(a_1, \dots, a_m; \{x_i, y_i, \sigma_{y_i}\}) = \sum_{i=1}^n \left[\frac{1}{\sigma_{y_i}^2} (y_i - y(x_i))^2 \right]$$

W minimum musi być spełniony układ m równań o postaci:

$$\begin{aligned}\frac{\partial \chi^2}{\partial a_l} &= \frac{\partial}{\partial a_l} \sum_{i=1}^n \left[\frac{1}{\sigma_{y_i}^2} (y_i - y(x_i))^2 \right] = 0 \\ &= -2 \sum_{i=1}^n \left[\frac{1}{\sigma_{y_i}^2} (y_i - y(x_i)) \frac{\partial y(x_i)}{\partial a_l} \right]\end{aligned}$$

Zachowanie się funkcji χ^2 w pobliżu minimum

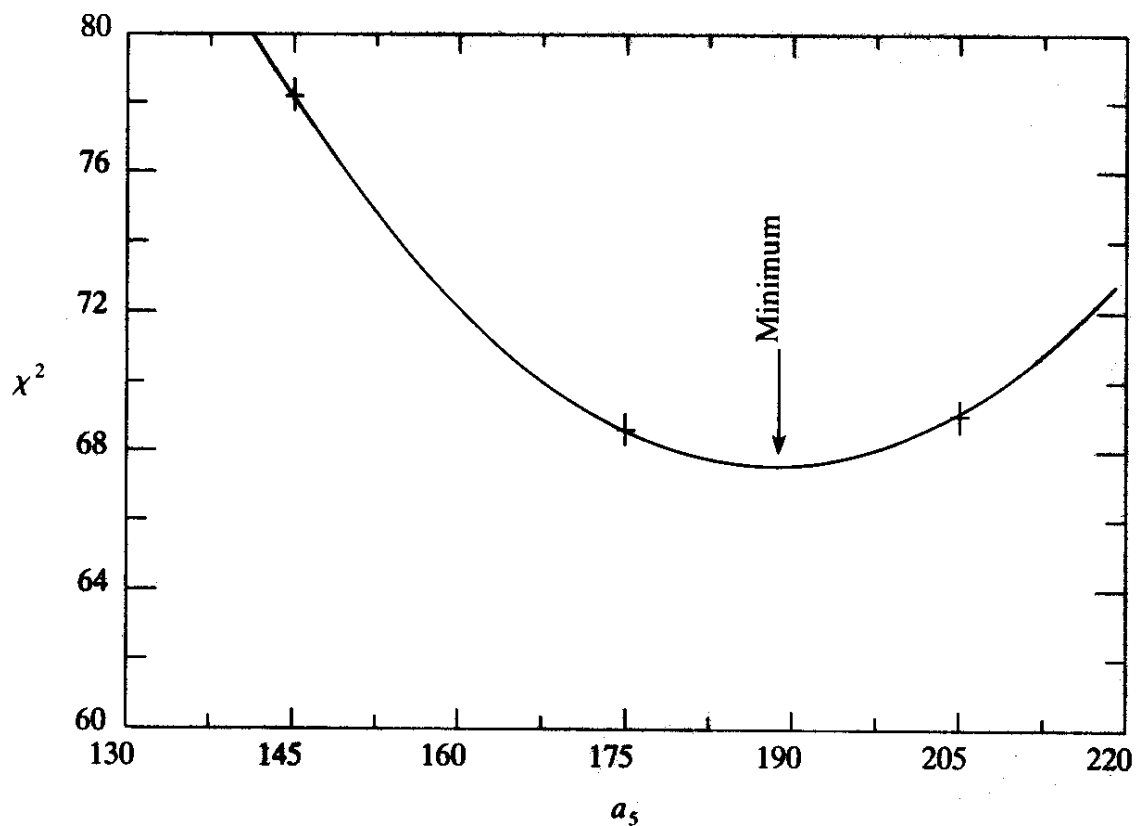


Dla uproszczenia rozważymy zależność χ^2 od tylko jednego parametru a_1 . χ^2 jest funkcją analityczną (jeżeli tylko nasz model wyraża się funkcją analityczną) i można ją rozwinąć w szereg potęgowy wokół minimum przypadającego dla wartości a'_1 :

$$\chi^2(a_1) = \chi^2(a'_1) + \frac{\partial \chi^2}{\partial a_1}(a_1 - a'_1) + \frac{1}{2} \frac{\partial^2 \chi^2}{\partial a_1^2}(a_1 - a'_1)^2 + \dots$$

Dostatecznie blisko minimum możemy pominąć wyrazy wyższych rzędów i wykorzystać fakt, że w samym minimum znika pierwsza pochodna.

$$\chi^2(a_1) \cong \chi^2(a'_1) + \frac{1}{2} \frac{\partial^2 \chi^2}{\partial a_1^2}(a_1 - a'_1)^2$$



Z drugiej strony, jeżeli rozważymy zachowanie się funkcji wiarygodności wokół wartości a'_l , to dla dostatecznie dużej liczby n punktów będzie ono zbliżone do funkcji rozkładu normalnego Gaussa

$$P(a_l) = A e^{-(a_l - a'_l)^2 / 2\sigma_l^2},$$

a wartość χ^2 można alternatywnie przedstawić jako

$$\chi^2 = -2 \ln[P(a_1, \dots, a_m)] + 2 \sum \ln(\sigma_i \sqrt{2\pi})$$

Po podstawieniu otrzymujemy równanie

$$\chi^2 = \frac{(a_l - a'_l)^2}{\sigma_l^2} + C$$

przedstawiające zależność χ^2 w pobliżu minimum od zmian wartości parametru a_l . Wynika z niego, że w pobliżu minimum funkcja χ^2 zachowuje się jak funkcja kwadratowa (zgodnie z tym co niezależnie wynika z rozwinięcia w szereg potęgowy) oraz, co ważne z praktycznego punktu widzenia, zmiana wartości parametru a_l o $\pm \sigma_{a_l}$ w stosunku do optymalnej a'_l powoduje wzrost χ^2 o 1.

Przez porównanie możemy jeszcze zapisać

$$\frac{1}{2} \frac{\partial^2 \chi^2}{\partial a_l^2} = \frac{1}{\sigma_l^2}$$

otrzymując związek między wariancją parametru a_l a krzywizną funkcji χ^2 w minimum:

$$\sigma_l^2 = 2 \left(\frac{\partial^2 \chi^2}{\partial a_l^2} \right)^{-1}.$$

Uzasadnienie na przykładzie wyznaczania wartości średniej.

$$P(a; \{y_i, \sigma\}) = \prod_{i=1}^n \left(\frac{1}{\sigma \sqrt{2\pi}} \right) \cdot \exp \left[-\frac{1}{2} \sum_{i=1}^n \frac{(y_i - a)^2}{\sigma^2} \right]$$

Zauważmy, że

$$(y_i - a)^2 = (a' + \varepsilon_i - a)^2 = (a' - a)^2 + 2(a' - a)\varepsilon_i + \varepsilon_i^2$$

gdzie a' jest optymalną wartością dopasowywanego parametru (czyli średnią) a ε_i ma rozkład $N(0, \sigma)$.

Wtedy

$$\begin{aligned} \sum_{i=1}^n \frac{(y_i - a)^2}{\sigma^2} &= \sum_{i=1}^n \frac{1}{\sigma^2} \left[(a' - a)^2 + 2(a' - a)\varepsilon_i + \varepsilon_i^2 \right] \\ &= \frac{n}{\sigma^2} (a' - a)^2 + \frac{2}{\sigma^2} (a' - a) \sum \varepsilon_i + \frac{1}{\sigma^2} \sum \varepsilon_i^2 \end{aligned}$$

Dla dużych wartości n zachodzi

$$\sum \varepsilon_i \cong 0, \quad \sum \varepsilon_i^2 \cong n\sigma^2$$

oraz dodatkowo

$$\sigma(a') = \frac{\sigma}{\sqrt{n}}$$

Możemy zatem zapisać

$$P(a; \{y_i, \sigma\}) \cong \prod_{i=1}^n \left(\frac{1}{\sigma \sqrt{2\pi}} \right) \cdot \exp \left[-\frac{1}{2} n \right] \cdot \exp \left[-\frac{(a' - a)^2}{2\sigma_{a'}^2} \right]$$

czyli rzeczywiście

$$P(a) = A \cdot \exp \left[-\frac{(a - a')^2}{2\sigma_{a'}^2} \right]$$

Metody minimalizacji polegają na przeszukiwaniu przestrzeni parametrów $\{a_1, a_2, a_3\}$ i znalezieniu w niej punktu, w którym (z dostateczną dokładnością) wartość funkcji χ^2 osiąga globalne minimum, albo na przybliżonym rozwiązaniu układu równań nieliniowych.

Metoda siatki (niezależnych kierunków) jest bardzo prostą metodą, przydatną w sytuacji, gdy zależność χ^2 od każdego z parametrów a_j słabo zależy od wartości pozostałych parametrów. Funkcję χ^2 minimalizujemy po kolei względem każdego parametru oddzielnie, powtarzając operacje do osiągnięcia zanedbywalnie małych zmian χ^2 .

1. Wybieramy początkowe wartości $\{a_j\}_0$, to znaczy punkt początkowy w przestrzeni parametrów i wartości kroku dla każdego parametru Δa_j oraz obliczamy χ_0^2 w punkcie początkowym.
2. Powiększmy parametr a_1 o $\pm \Delta a_1$ i obliczamy χ_1^2 (znak dobrany tak, żeby zmniejszyć χ^2).
3. Powtarzamy krok 2. do momentu, kiedy χ^2 przestaje się zmniejszać. Wzrost χ^2 oznacza przekroczenie „dna doliny” i wspięcie się po drugiej jej stronie.
4. Do ostatnich trzech położzeń (i odpowiadających im wartościom χ^2) obejmujących minimum w kierunku „marszu” dopasowujemy parabolę. Minimum dopasowanej paraboli przyjmujemy jako punkt początkowy dla minimalizacji względem kolejnego parametru.
5. Powtarzamy kroki 2., 3. i 4. minimalizując χ^2 po kolei dla każdego parametru.
6. Całą procedurę powtarzamy, aż do zlokalizowania minimum z pożądaną dokładnością.

Metoda największego spadku (gradientu) wykorzystuje fakt, że wektor gradientu funkcji χ^2

$$\vec{\nabla}\chi^2 = \sum_{j=1}^m \left[\frac{\partial\chi^2}{\partial a_j} \vec{e}_j \right]$$

wskazuje kierunek największego wzrostu. Kierunek przeciwny do zwrotu wektora gradientu jest kierunkiem największego spadku. Zatem zmiana początkowych wartości parametrów

$$\mathbf{a}_{n+1} = \mathbf{a}_n - \text{const} \cdot \vec{\nabla}\chi^2(\mathbf{a}_n)$$

gwarantuje zmniejszenie wartości χ^2 . Problemem jest odpowiednie dobranie wielkości mnożnika oraz fakt, że blisko minimum gradient ma bardzo małe wartości i staje się praktycznie bezużyteczny (w minimum gradient znika).

Metoda rozwinięcia funkcji χ^2 polega na zastąpieniu dokładnej postaci tej funkcji (i odpowiadającej jej hiperpowierzchni w m -wymiarowej przestrzeni parametrów) jej rozwinięciem w szereg z wyrazami do drugiego stopnia włącznie (czyli paraboloidalną powierzchnią drugiego stopnia).

$$\chi^2 \cong \chi_0^2 + \sum_{j=1}^m \left[\frac{\partial\chi_0^2}{\partial a_j} \delta a_j \right] + \frac{1}{2} \sum_{k=1}^m \sum_{j=1}^m \left[\frac{\partial^2\chi_0^2}{\partial a_j \partial a_k} \delta a_j \delta a_k \right]$$

Wartości χ^2 i jej pochodnych po prawej stronie są obliczone w punkcie początkowym \mathbf{a}_0 , przyrosty δa_j są zdefiniowane jako:

$$\delta a_j = a_j - a_{j0}$$

Wartość χ^2 po lewej stronie jest funkcją m przyrostów parametrów δa_j .

Minimum tej przybliżonej funkcji χ^2 (powierzchni paraboloidalnej) wyznaczają warunki

$$\frac{\partial\chi_0^2}{\partial(\delta a_k)} = \frac{\partial\chi_0^2}{\partial a_k} + \sum_{j=1}^m \left[\frac{\partial^2\chi_0^2}{\partial a_j \partial a_k} \delta a_j \right] = 0, \quad k = 1, \dots, m,$$

które tworzą układ m równań liniowych względem przyrostów δa_j .

Jeżeli oznaczymy

$$\beta_k = -\frac{1}{2} \frac{\partial \chi_0^2}{\partial a_k} \quad \text{i} \quad \alpha_{jk} = \alpha_{kj} = \frac{1}{2} \frac{\partial^2 \chi_0^2}{\partial a_j \partial a_k},$$

to układ równań możemy zapisać w postaci układu macierzowego.

$$\boldsymbol{\beta} = \boldsymbol{\delta} \mathbf{a} \boldsymbol{\alpha}$$

Przykład.

W pracowni fizycznej wyznacza się stałe zaniku izotopów promieniotwórczych srebra aktywowanego strumieniem neutronów zliczając impulsy licznika G-M rejestrującego promieniowanie emitowane z płytki srebrnej. Na podstawie prawa zaniku promieniotwórczego przewidujemy, że liczba zliczeń w kolejnych jednakowych odstępach czasu będzie się zmieniała jak funkcja:

$$y(x) = a_1 + a_2 \cdot e^{-a_3 \cdot x} + a_4 \cdot e^{-a_5 \cdot x}$$

gdzie $y(x)$ jest oczekiwaną liczbą impulsów zarejestrowanych w kolejnym odstępie czasu, po czasie x od momentu rozpoczęcia pomiarów. a_1 - ma sens biegu własnego licznika (tła), a_2, a_4 - początkowej liczby zliczeń od każdego izotopu, a_3, a_5 - szukane stałe zaniku. Ze względu na fakt, że funkcja $y(x)$ zależy nieliniowo od parametrów a_3 i a_5 , dopasowanie tej funkcji do danych doświadczalnych jest zagadnieniem nieliniowym.

Dla uproszczenia rachunków uwzględnimy tylko jeden izotop.

Funkcja χ^2 ma następującą postać

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

Warunek jaki muszą spełniać wartości parametrów w minimum da się zapisać jako układ trzech równań:

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

Po uproszczeniach i przy założeniu, że $a_2 \neq 0$ otrzymujemy

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

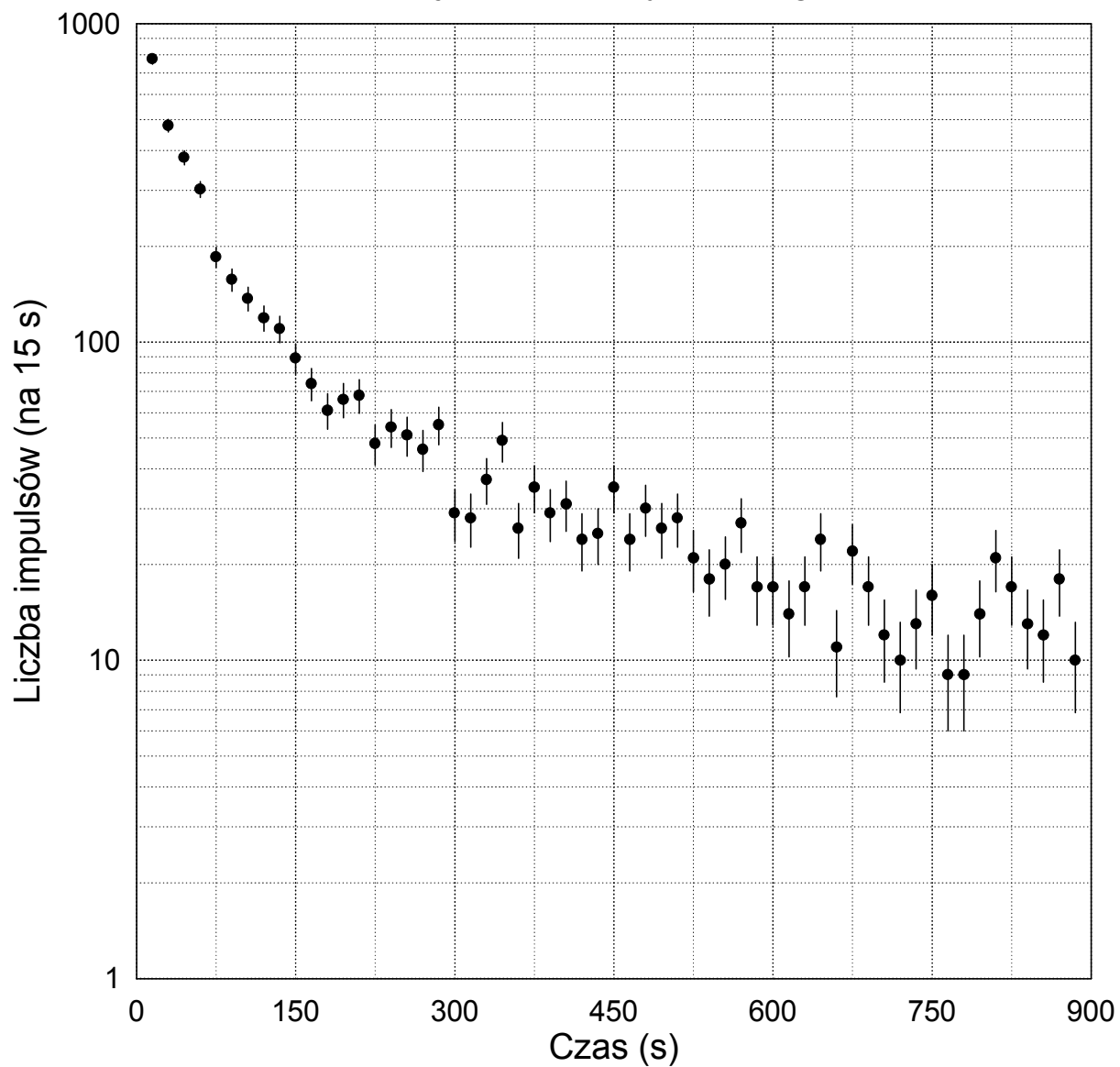
układ trzech równań nieliniowych względem parametru a_3 .

Dane liczbowe do przykładu w wykładzie 10.

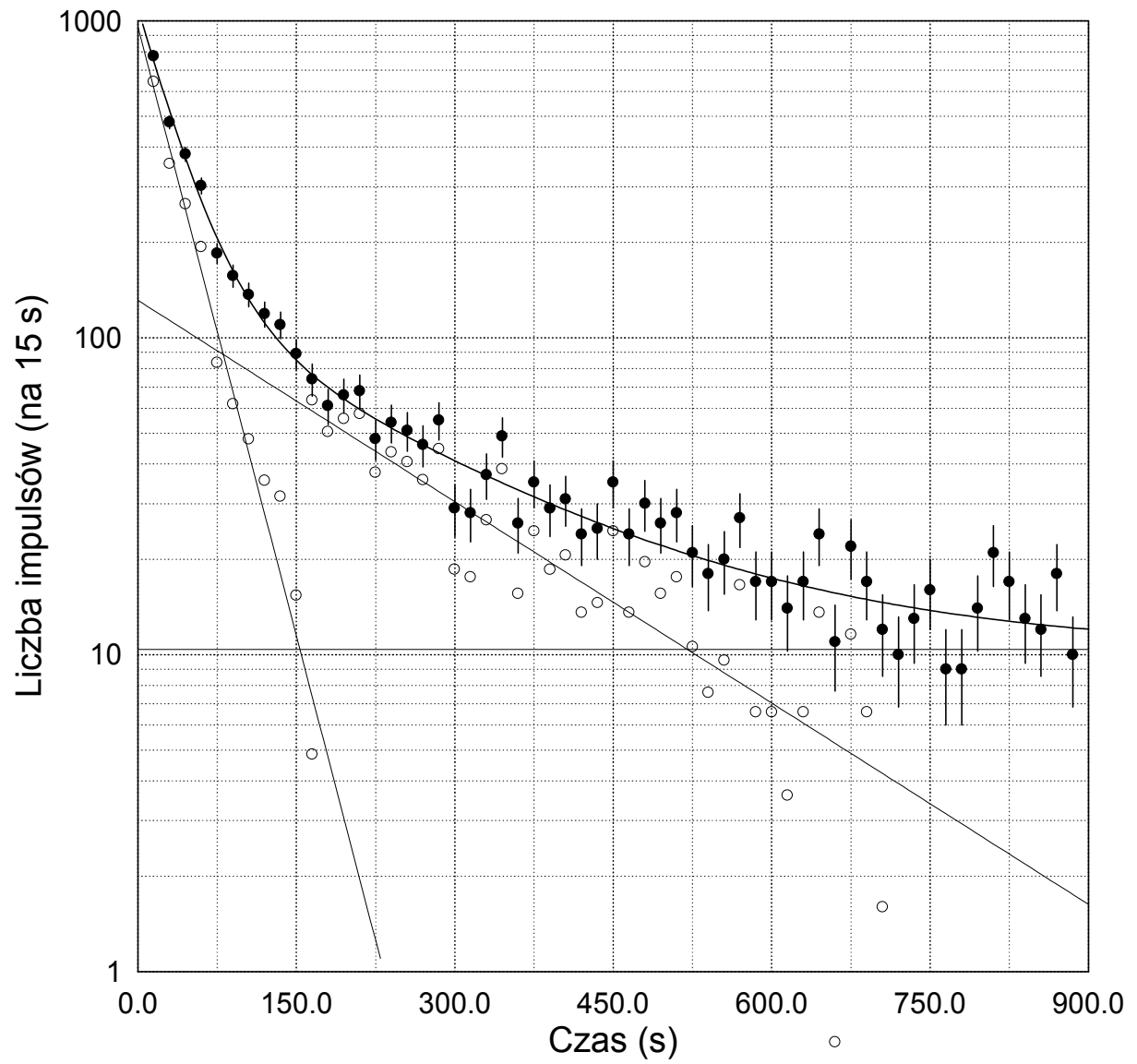
x_i - czas, s	y_i - liczba impulsów	$\sigma(y_i)$ - niepewność
15	775	27,84
30	479	21,89
45	380	19,49
60	302	17,38
75	185	13,60
90	157	12,53
105	137	11,70
120	119	10,91
135	110	10,49
150	89	9,43
165	74	8,60
180	61	7,81
195	66	8,12
210	68	8,25
225	48	6,93
240	54	7,35
255	51	7,14
270	46	6,78
285	55	7,42
300	29	5,39
315	28	5,29
330	37	6,08
345	49	7,00
360	26	5,10
375	35	5,92
390	29	5,39
405	31	5,57
420	24	4,90
435	25	5,00
450	35	5,92
465	24	4,90
480	30	5,48
495	26	5,10
510	28	5,29
525	21	4,58
540	18	4,24
555	20	4,47
570	27	5,20
585	17	4,12
600	17	4,12
615	14	3,74
630	17	4,12
645	24	4,90
660	11	3,32

675	22	4,69
690	17	4,12
705	12	3,46
720	10	3,16
735	13	3,61
750	16	4,00
765	9	3,00
780	9	3,00
795	14	3,74
810	21	4,58
825	17	4,12
840	13	3,61
855	12	3,46
870	18	4,24
885	10	3,16

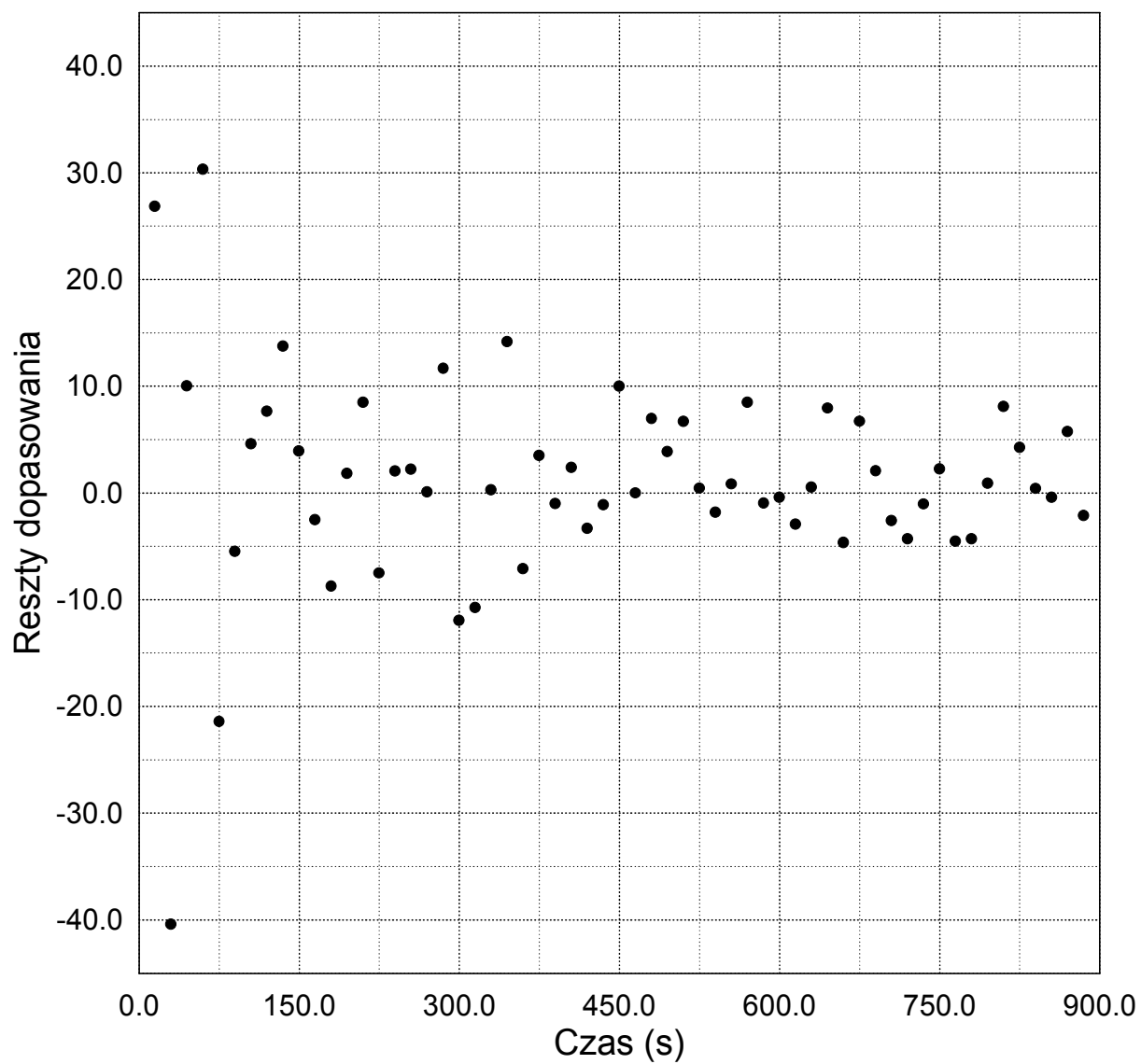
Liczba impulsów rejestrowanych w czasie obserwacji rozpadu aktywowanego srebra



Dopasowanie dwóch eksponent i tła do wyników pomiarów aktywności



Rozrzut reszt dopasowania wokół zera



Rozrzut unormowanych reszty dopasowania wokół zera

