SPRAWOZDANIE NUM8

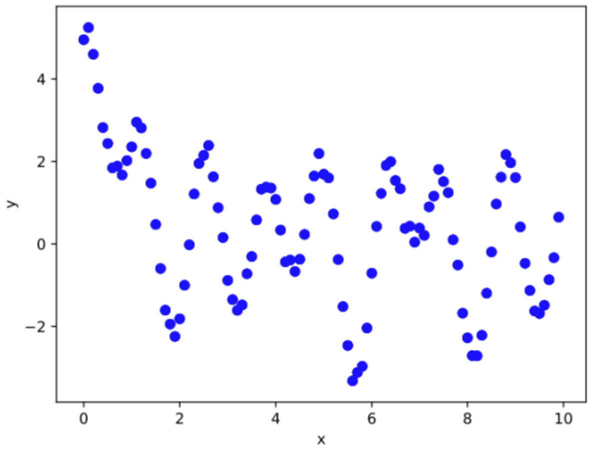
JAKUB KRĘCISZ

Treść zadania

Zadany jest zbiór punktów zilustrowany poniżej, dwie liczby w każdym wierszu to współrzędne x i y. Punkty te modelujemy za pomocą funkcji $F(x) = a \cdot \sin(2x) + b \cdot \sin(3x) + c \cdot \cos(5x) + d \cdot \exp(-x)$.

(a) Znajdź wartości współczynników a-d które najlepiej opisują te dane w sensie metody najmniejszych kwadratów. Rezultat przedstaw graficznie. Rozwiązując to zadanie nie można korzystać z procedur bibliotecznych służących do aproksymacji. Poza tym, użycie procedur z zakresu algebry liniowej jest dozwolone.

(b) Zaproponuj inną funkcję $G(x)$ (która zależy od kilku parametrów) i wygeneruj zbiór punktów w postaci $(x, G(x) + \delta y)$, gdzie δy to losowe zaburzenia. Powtórz dopasowanie z pkt. (a) dla swoich danych i sprawdź, czy udało się odtworzyć wartości ustalonych wcześniej parametrów. Poeksperymentuj zmieniając ilość wygenerowanych punktów i wielkość zaburzeń.



Rysunek 1: Dane do zadania numerycznego NUM8.

Omówienie zadania

Zacznijmy od tego, czym w zasadzie jest metoda najmniejszych kwadratów. Jest to metoda służąca do określania przybliżenia współczynników funkcji, tak aby sama funkcja jak najlepiej pasowała do danych (zazwyczaj zbiór punktów (x, y)). Wiedząc, że funkcja jest znana co do swojego kształtu, a tylko nieznane są jej współczynniki, możemy za pomocą metody najmniejszych kwadratów odnaleźć możliwie najbliższe wartości współczynników, które będą naszą funkcję możliwie najbliżej przybliżać do przebiegania przez podane punkty. Polega ona na minimalizacji sumy kwadratów różnic pomiędzy rzeczywistymi wartościami y a przewidywanymi wartościami y dla danego zbioru punktów.

Zacznijmy od przyjęcia, że mamy N par punktów takich, że:

$$\{(x_i, y_i)\}_{i=1}^{N}$$

Gdzie x_i jest dokładną wartością argumentu a y_i jest zmierzoną wartością funkcji, które mogą być obarczone błędami pomiarowymi.

Każdej zmierzonej wartości y_i odpowiada wartość teoretyczna $\tilde{y}_i$ jaką zmienna powinna przybrać dla danej wartości zmiennej x. Przyjmujemy, że wartość teoretyczna jest kombinacją liniową pewnych znanych funkcji:

$$\tilde{y}_i = a_1 \cdot f_1(x_i) + a_2 \cdot f_2(x_i) + \cdots + a_s \cdot f_s(x_i)$$

W naszym przypadku dla funkcji z podpunktu (a):

$$\tilde{y}_i = a_1 \cdot \sin(2x_i) + a_2 \cdot \sin(3x_i) + a_3 \cdot \cos(5x_i) + a_4 \cdot \exp(-x_i)$$

Zespół wszystkich wartości teoretycznych, możemy zatem przedstawić jako:

$$\tilde{y} = Ap$$

Gdzie dla przykładu podpunkt (a) to:

$$A = \begin{bmatrix} \sin(2x_1) & \sin(3x_1) & \cos(5x_1) & \exp(-x_1) \\ \sin(2x_2) & \sin(3x_2) & \cos(5x_2) & \exp(-x_2) \\ \vdots & \vdots & \vdots & \vdots \\ \sin(2x_N) & \sin(3x_N) & \cos(5x_N) & \exp(-x_N) \end{bmatrix}$$

$$p = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix}$$

Czyli:

$$\tilde{y} = \begin{bmatrix} \sin(2x_1) & \sin(3x_1) & \cos(5x_1) & \exp(-x_1) \\ \sin(2x_2) & \sin(3x_2) & \cos(5x_2) & \exp(-x_2) \\ \vdots & \vdots & \vdots & \vdots \\ \sin(2x_N) & \sin(3x_N) & \cos(5x_N) & \exp(-x_N) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix}$$

Do policzenia naszego problemu posłużymy się wzorem dla metody najmniejszych kwadratów. Gdzie b to wektor wartości przybliżonych y.

$$Ap = b \leftrightarrow A^T Ap = A^T b$$

Co nam daje dokładnie:

$$\begin{bmatrix} \sin(2x_1) & \sin(3x_1) & \cos(5x_1) & \exp(-x_1) \\ \sin(2x_2) & \sin(3x_2) & \cos(5x_2) & \exp(-x_2) \\ \vdots & \vdots & \vdots & \vdots \\ \sin(2x_N) & \sin(3x_N) & \cos(5x_N) & \exp(-x_N) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

Dzięki SVD (Singular Value Decomposition), jesteśmy w stanie obliczyć odwrotność naszej macierzy $A^T A$, dzięki której możemy nasze równanie $A^T Ap = A^T b$ przekształcić na wzór na minimalizację sumy kwadratu błędów, wtedy nasze równanie będzie wyglądało:

$$A^T Ap = A^T b \leftrightarrow p = (A^T A)^{-1} A^T b$$

Dzięki temu, jesteśmy w stanie wprost wyznaczyć przybliżenie naszych współczynników dla podanej funkcji. Nasze współczynniki to kolejne wartości w wektorze p.

Bardzo podobnie podchodzimy do podpunktu (b). W którym mamy zaproponować inną funkcję G(x). Jedyną dodatkową rzeczą w tym przypadku jest wyznaczenie zbioru punktów z nadanym losowym zaburzeniem dla każdej wartości y. Zbiór naszych punktów dla podpunktu (b) będzie miał postać:

$$\begin{bmatrix} x_1 & G(x_1) + \delta y_1 \\ x_2 & G(x_2) + \delta y_2 \\ \vdots & \vdots \\ x_N & G(x_N) + \delta y_N \end{bmatrix}$$

Gdzie pierwsza kolumna to zbiór wartości x, druga kolumna to zbiór wartości y, gdzie dla każdej wartości dodajemy losowe zaburzenie, które oznaczamy jako δy_i.

Uruchomienie programu

Do uruchomienia programu wykorzystamy Makefile:

Aby uruchomić program dla podpunktu (a) i wyświetlić wykres, wystarczy użyć polecenia:

make a_show

Aby uruchomić program dla podpunktu (a) i zapisać wykres w postaci pliku svg, wystarczy użyć polecenia:

make a_save

Aby uruchomić program dla podpunktu (b) i wyświetlić wykres, wystarczy użyć polecenia:

make b_show

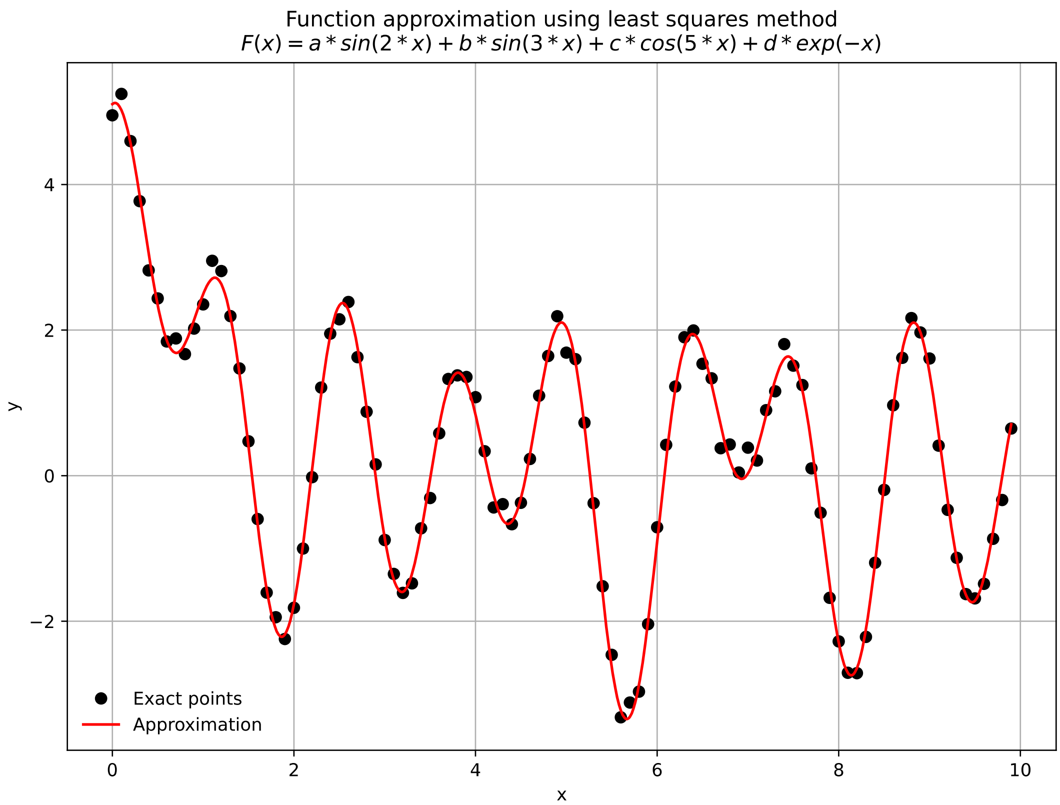
Aby uruchomić program dla podpunktu (b) i zapisać wykres w postaci pliku svg, wystarczy użyć polecenia:

make b_save

Wyniki

(Przyjmijmy oznaczenie d, oznaczające wektor dokładnych współczynników)

Podpunkt (a)

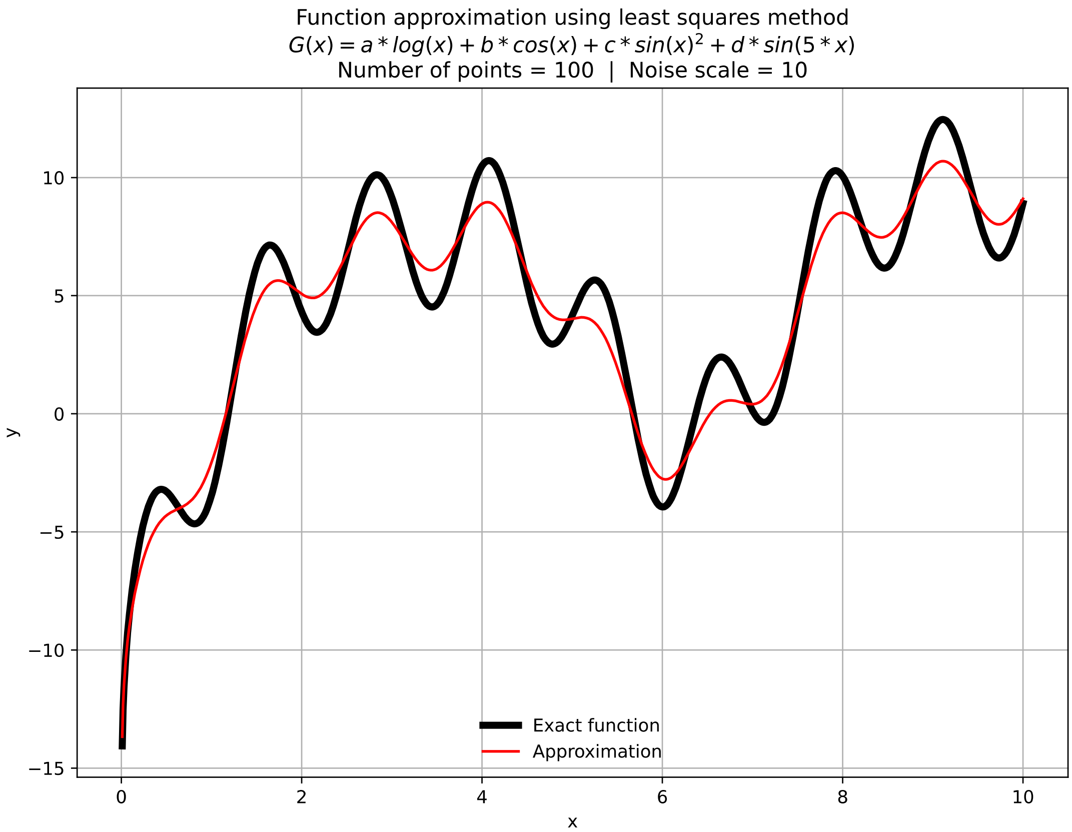


Funkcja: $F(x) = a \cdot \sin(2x) + b \cdot \sin(3x) + c \cdot \cos(5x) + \exp\,(-x)$

Znalezione współczynniki:

$$p = \begin{bmatrix} 0.66767124 \\ 1.07293276 \\ 1.69693956 \\ 3.40636597 \end{bmatrix}$$

Podpunkt (b):



Funkcja: $G(x) = a \cdot log(x) + b \cdot cos(x) + c \cdot sin(x)^\wedge 2 + d \cdot sin(5 * x)$

Liczba punktów: 100

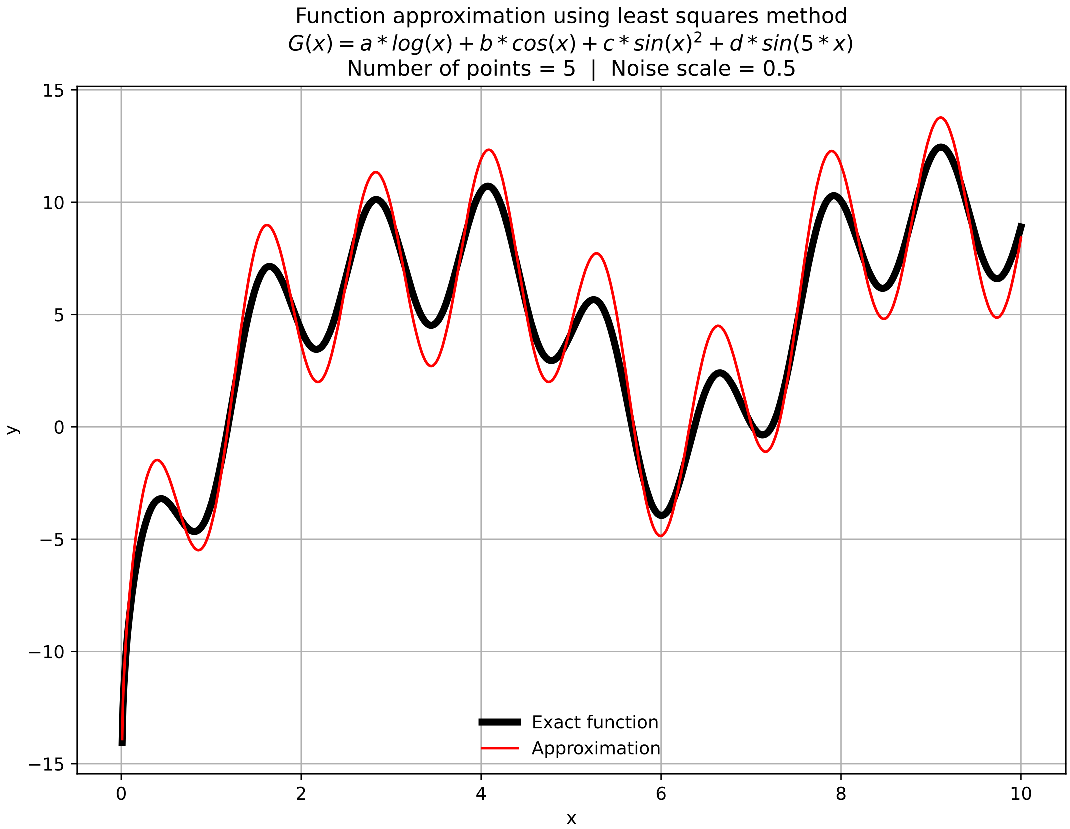
Skala szumów: 10

Dokładne współczynniki:

$$d = \begin{bmatrix} 2 \\ -5 \\ 3 \\ 3 \end{bmatrix}$$

Znalezione współczynniki to:

$$p = \begin{bmatrix} 2.34195183 \\ -6.89946845 \\ 2.48936023 \\ 4.29385315 \end{bmatrix}$$



Funkcja: $G(x) = a \cdot log(x) + b \cdot cos(x) + c \cdot sin(x)\text{^}2 + d \cdot sin(5 * x)$

Liczba punktów: 5

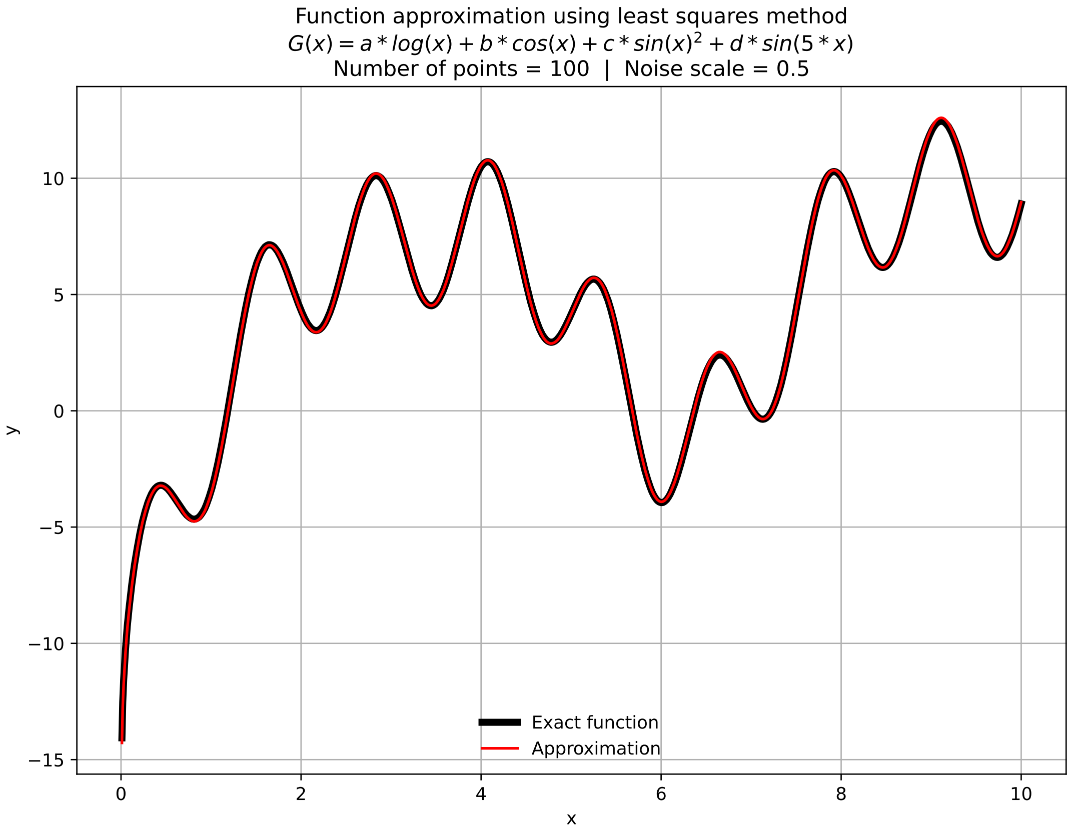
Skala szumów: 0.5

Dokładne współczynniki:

$$d = \begin{bmatrix} 2 \\ -5 \\ 3 \\ 3 \end{bmatrix}$$

Znalezione współczynniki to:

$$p = \begin{bmatrix} 2.07823554 \\ -4.55641522 \\ 3.37831187 \\ 4.52272938 \end{bmatrix}$$



Funkcja: $G(x) = a \cdot log(x) + b \cdot cos(x) + c \cdot sin(x)^\wedge 2 + d \cdot sin(5 * x)$

Liczba punktów: 100

Skala szumów: 0.5

Dokładne współczynniki:

$$d = \begin{bmatrix} 2 \\ -5 \\ 3 \\ 3 \end{bmatrix}$$

Znalezione współczynniki to:

$$p = \begin{bmatrix} 2.00839135 \\ -4.99939452 \\ 2.98915365 \\ 3.00488222 \end{bmatrix}$$

Wnioski

Jak widać w wynikach, w punkcie (a) udało nam się odnaleźć współczynniki a-d które najlepiej będą opisywać tą funkcję w sensie metody najmniejszych kwadratów na podstawie tej ilości punktów którą nam dano.

Idąc dalej, w podpunkcie (b), eksperymentując, czyli zmieniając ilość wygenerowanych punktów i wielkość zaburzeń możemy zauważyć w szczególności, że gdy ilość wygenerowanych punktów będzie niewystarczająca, nasze przybliżenie współczynników funkcji będzie coraz gorsze. To samo się dzieje w przypadku, gdy wielkość zaburzeń będzie zbyt duża. Jednak, gdy podamy wystarczającą ilość punktów a punkty same w sobie będą całkiem dokładne (skala zaburzeń będzie odpowiednio mała) to nasze przybliżenie współczynników może być bardzo dokładne (nawet do 7 cyfr po przecinku dla dla przykładu 1000 punktów i skali zaburzeń równej 0.000001).

Wniosek jest prosty, wraz ze wzrostem ilości punktów i spadkiem skali zaburzeń, jesteśmy w stanie odnaleźć coraz to dokładniejsze wartości współczynników na podstawie metody najmniejszych kwadratów.