

## Spis treści

Przed wykładami   Algebra i geometria.....2	
Wykład 1   Zagadnienia wstępne .....2	
Proporcje błędów.....2	
Reguła sumacyjna Kahana .....3	
Wykład 2   Uwarunkowania problemów numerycznych .....3	
Norma wektora .....3	
Współczynnik uwarunkowania .....3	
Eliminacja Gaussa .....3	
Wybór elementu podstawowego ....3	
Wykład 3   Faktoryzacje .....4	
Faktoryzacja LU .....4	
Algorytm Doolittle'a .....4	
Algorytm Crouta .....4	
Algorytm Thomasa .....4	
Faktoryzacja Cholesky'ego.....4	
Faktoryzacja LDL.....4	
Macierze rzadkie, wypełnienie a faktoryzacja LU i Cholesky'ego .....4	
Faktoryzacja QR.....5	
Transformacja Householdera .....5	
Obroty Givensa.....5	
Wzór Shermana-Morrisona .....5	
Singular Value Decomposition.....5	
Nadokreślone układy równań.....5	
Wykład 4   Metody iteracyjne .....5	
Metoda Jacobiego .....5	
Metoda Gaussa-Seidela.....6	
Metoda gradientów sprzężonych ....6	
Conjugate Gradients (CG .....6	
Wykład 5   Numeryczne zagadnienie własne .....6	
Algorytm PageRank .....7	
Równanie charakterystyczne macierzy .....7	
Metoda potęgowa.....7	
Odwrotna metoda potęgowa .....7	
Transformacja podobieństwa.....7	
Ortogonalna transformacja podobieństwa .....7	
Algorytm diagonalizacji .....8	
Uwagi .....8	
Macierze Hermitowskie.....8	
Rezolwenta.....8	
Transformacja Mobiusa.....8	
Uogólnione zagadnienie własne .....8	
Wykład 6   Interpolacja .....8	
Interpolacja odcinkami liniowa.....8	

Interpolacja wielomianowa ..... 8	
Wzór interpolacyjny Lagrange'a..... 9	
Oscylacje Rungego ..... 9	
Błąd interpolacyjny ..... 9	
Współczynniki wielomianu interpolacyjnego ..... 9	
Interpolacja Hermite'a ..... 9	
Interpolacja za pomocą funkcji sklepanych (splajny)..... 9	
Splajny kubiczne ..... 9	
Naturalny splajn kubiczny ..... 9	
Funkcje sklepane Catmulla-Roma .. 10	
Splajny bikubiczne ..... 10	
Interpolacja za pomocą funkcji wymiennych ..... 10	
Algorytm Floatera i Hormanna..... 10	
Różniczkowanie numeryczne ..... 10	
Różniczkowanie splajnów ..... 10	
Wykład 7   Całkowanie numeryczne... 11	
Kwadratury ..... 11	
Metoda trapezów ..... 11	
Metoda Simpsona..... 11	
Metoda 3/8..... 11	
Metoda Milne'a ..... 11	
Złożony wzór trapezów ..... 11	
Ekstrapolacja Richardsona ..... 12	
Metoda Romberga..... 12	
Całkowanie po przedziałach nieskończonych ..... 12	
Kwadratury adaptacyjne ..... 12	
Całki wymiarowe ..... 12	
Kwadratury adaptacyjne w wielu wymiarach ..... 12	
Wykład 8   Rozwiązywanie równań algebraicznych ..... 13	
Metoda bisekcji ..... 13	
Metoda reguła fałsi..... 13	
Metoda siecznych ..... 13	
Interpolacja odwrotna ..... 13	
Kryterium stopu..... 13	
Metoda Newtona..... 13	
Tłumiona metoda Newtona ..... 14	
Metody wykorzystujące drugą pochodną..... 14	
Metoda Halleya ..... 14	
Układy równań algebraicznych ..... 14	
Wielowymiarowa metoda Newtona ..... 14	
Rozwiązywanie równań nieliniowych a minimalizacja ..... 14	
Metoda globalnie zbieżna ..... 14	

Wielowymiarowa metoda siecznych – metoda Broydena..... 15	
Wykład 9   Miejsca zerowe wielomianów ..... 15	
Algorytm Hornera ..... 15	
Przykład Wilkinsona ..... 15	
Metoda Laguerre'a..... 15	
Wykład 10   Minimalizacja funkcji jednej zmiennej..... 16	
Wstępna lokalizacja minimum..... 16	
Metoda złotego podziału ..... 17	
Metoda paraboliczna i metoda Brenta ..... 17	
Wykorzystanie pochodnych ..... 17	
Wykorzystanie interpolacji Hermite'a ..... 17	
Wykład 11   Minimalizacja funkcji wielu zmiennych..... 17	
Metoda najszybszego spadku (steepest descent, gradient descent) ..... 18	
Stochastic Gradient Descent ..... 18	
Zastosowanie metody Newtona .... 18	
Dodatnia określoność hessjanu .... 18	
Metoda Levenberga-Marquardta .. 18	
Strategia precyzyjnej minimalizacji wielowymiarowej ..... 18	
Funkcja Rosenbrocka..... 19	
Metoda gradientów sprzężonych .. 19	
Metoda zmiennej metryki ..... 19	
Metoda Powella ..... 19	
Wykład 12   Uwagi o minimalizacji globalnej ..... 19	
Algorytm Monte Carlo..... 19	
Algorytmy genetyczne..... 19	
Paralelizacja ..... 20	
Particle Swam Optimization ..... 20	
Wykład 13   Aproksymacja i zagadnienie najmniejszych kwadratów ..... 20	
Liniowe zagadnienie najmniejszych kwadratów ..... 20	
Błędy pomiarowe ..... 20	
Metoda najmniejszych kwadratów 20	
Forma kwadratowa estymatorów . 21	
Minimum formy kwadratowej..... 21	
Nadokreślony układ równań ..... 21	
Pomiary nieskorelowane..... 21	
Kryterium Akaike..... 21	
Nieliniowe zagadnienie najmniejszych kwadratów ..... 21	
Pseudolinearyzacja..... 21	
Wykład 14   Przybliżenia Pade ..... 22	

## Przed wykładami | Algebra i geometria

Macierz rzadka – wtedy gdy liczba niezerowych elementów rośnie wolniej niż  $N^2$

Macierz osobliwa – macierz o wyznaczniku nieodwracalnym (zerowym)

Macierz diagonalna – macierz która posiada elementy tylko na diagonalu.

Macierz diagonalizowalna (normalna) – macierz która ma  $N$  niezależnych linowo wektorów własnych.

Macierz trójkątna – macierz posiadająca elementy tylko nad/pod diagonalą.

Macierz ortogonalna – macierz kwadratowa  $A$  spełniająca własność:  $A^T \cdot A = A \cdot A^T = I$

Macierz hermitowska (samosprężona) – macierz kwadratowa  $A$  równa swojemu sprzężeniu hermitowskiemu:  $H^\dagger = H$

Sprzężenie hermitowskie – złożenie operacji transpozycji i sprzężenia zespolonego dokonane na macierzy w ogólności zespolonej. (Sprzężenie liczby zespolonej:  $z = a + bi \rightarrow \bar{z} = a - bi$ )

Macierz unitarna – macierz kwadratowa  $U$  o elementach zespolonych spełniająca własność:  $U^\dagger U = U U^\dagger = I$

Macierz Vandermonda – jest to macierz reprezentująca układ wielomianów, jej pierwszą kolumną są jedynki, kolejną  $x^1$  itd. aż do  $n-1$  potęgi  $x$ .

Macierz Hessenberga – macierz trójkątna górna z jedną dodatkową diagonalą pod główną.

Gradient to pole wektorowe wskazujące kierunki najszybszych wzrostów wartości danego pola skalarne w poszczególnych punktach.

Hessjan jest to macierz drugich pochodnych.

Podstawowe twierdzenie Algebry – wielomian stopnia  $n$  ma na płaszczyźnie zespolonej dokładnie  $n$  pierwiastków przy czym pierwiastki wielokrotne liczą się z ich krotnościami.

## Wykład 1 | Zagadnienia wstępne

W sposób ścisły w komputerze można reprezentować liczby całkowite (z pewnego zakresu) oraz liczby wymierne posiadające skończone rozwinięcie binarne (z pewnego zakresu). Wszystkie inne liczby można reprezentować tylko w sposób przybliżony – są one zatem obarczone pewnym błędem nazywanym błędem zaokrąglenia.

Liczba  $x$  jest poprawnie zaokrąglana na  $i$ -tej pozycji do liczby, którą oznaczamy  $x^{(i)}$ , jeśli błąd zaokrąglenia  $\varepsilon$  jest taki że  $|\varepsilon| = |x - x^{(i)}| \leq \frac{1}{2} \cdot 10^i$ . Każda cyfra poprawnie zaokrąglonej liczby, począwszy od pierwszej cyfry różnej od zera jest znacząca. Liczba cyfr znaczących jest pewną miarą błędu zaokrąglenia.

### Proporcje błędu

- Przy dodawaniu błąd się sumuje
- Przy mnożeniu błąd się mnoży, może to powodować znaczny wzrost błędu
- Przy dzieleniu błąd się mnoży, przy dzieleniu przez b. małe liczby błąd rośnie ekstremalnie

Dodawanie na komputerze nie zawsze jest łączne. Operujemy na arytmetyce ze skończoną dokładnością!

**Reguła sumacyjna Kahana** to algorytm, w którym trzyma się dodatkową zmienną akumulującą małe błędy. Dzięki temu błąd w najgorszym przypadku praktycznie nie zależy od  $n$ . Jest to po prostu iteracyjne sumowanie kolejnych liczb i odejmowanie błędu powstałego w poprzedniej iteracji. Bardzo ważne jest kolejność operacji w wyznaczaniu błędu do odjęcia była zachowana. Inne podejście do sumowania liczb to posortowanie ich i sumowanie w kolejności od najmniejszej do największej. Jeżeli liczby mają różne znaki to rozbijamy je na dwie zbiory: dodatnie i ujemne i sortujemy & sumujemy oba z osobna na końcu dodając obie sumy.

## Wykład 2 | Uwarunkowania problemów numerycznych

Zagadnienie obliczenia czegoś numerycznie jest dobrze uwarunkowane gdy niewielkie względne zmiany danych na wejściu dają niewielkie względne zmiany rozwiązania. W przeciwnym wypadku są to zagadnienia źle uwarunkowane.

**Norma wektora** jest to uogólnienie pojęcia wartości bezwzględnej na przypadek wektorów.

- Norma taksówkowa  $\|x\|_1 = |x_1| + |x_2| + \dots + |x_n|$
- Norma Euklidesowa  $\|x\|_2 = \sqrt{x^T x} = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$
- Norma maximum (worst offender)  $\|x\|_\infty = \max_{i=1, \dots, n} |x_i|$

Domyślnie jako norma wektorowa używana jest norma Euklidesowa, chyba że zaznaczymy inaczej.

**Współczynnik uwarunkowania** mówi jak bardzo błąd względny wyniku obliczeń „przekracza” błąd względny samej różnicy przybliżenia i wartości dokładnej. Służy do wskazania jak bardzo źle jest uwarunkowane zagadnienie. Im większy ten współczynnik tym gorzej. Współczynnik uwarunkowania macierzy symetrycznych rzeczywistych wynosi  $\kappa = \frac{\max_i |\lambda_i|}{\min_i |\lambda_i|}$

Układ równań  $Ax = b$  ma jednoznaczne rozwiązanie wtedy i tylko wtedy gdy  $\det A \neq 0$

Do rozwiązywania układów równań nie używamy wzorów Cramera bo jest to b. drogie numerycznie.

Równania można dodać stronami po pomnożeniu przez dowolną stałą różną od zera, można przestawiać kolejność (permutacja kolumn i wierszy).

**Eliminacja Gaussa** czyli odejmowanie & dzielenie kolejno współczynników sprowadzi nam macierz do postaci trójkątnej którą można rozwiązywać za pomocą backwards substitution ale jest to bardzo nieefektywne bo wynosi  $O(N^3)$  i dodatkowo algorytm się zawali gdy będziemy musieli dzielić przez zero (można tego uniknąć mądrze przestawiając kolumny/wiersze itp.)

**Wybór elementu podstawowego** sprowadza się do tego, by unikać dzielenia przez bardzo małe liczby, dlatego przy każdej „iteracji” szukamy najbardziej odpowiedniego elementu (czyli największego co do modułu) i przestawiamy kolumny/wiersze w ten sposób aby to przez niego dzielić. Koszt wykonania takich wyborów elementu podstawowego w całej macierzy wnosi  $O(N^2)$  i opłaca się go zawsze robić przy eliminacji Gaussa. Pełny wybór elementu podstawowego polega na uwzględnianiu zawsze wszystkich współczynników a nie tylko tych w aktualnie operowanej kolumnie. Koszt tego wynosi  $O(N^3)$  więc porównywalnie z Gaussem.

Do skutecznego przeprowadzenia eliminacji Gaussa potrzebna jest znajomość kolumny wyrazów wolnych bo one też są przekształcane i permutowane w czasie eliminacji.

Nigdy nie konstruujemy macierzy odwrotnej. Jest to zbyt drogie bo aż  $O(2N^3)$

## Wykład 3 | Faktoryzacje

Faktoryzacja to przekształcenie macierzy równań w iloczyn macierzy.

### Faktoryzacja LU

Polega na znalezieniu faktoryzacji  $A = L \cdot U$  gdzie  $L$  jest macierzą trójkątną dolną z jedynkami na diagonalu a  $U$  trójkątną górną.  $Ax = LUx = b$  Rozwiązujemy dwa równania:

$$\begin{cases} Ux = y \\ Ly = b \end{cases}$$

Koszt ich rozwiązania wynosi  $O(2N^2)$

### Algorytm Doolittle'a

Polega na wyznaczaniu faktoryzacji LU przez liczenie kolejnych elementów macierzy  $L$  i  $U$  w dobrej kolejności. Złożoność tego algorytmu wynosi  $O(N^3)$ . Wykonanie tej faktoryzacji nie wymaga dodatkowej pamięci bo jest wykonywana w miejscu (zamieniamy elementy na nowe). Opłaca się przeprowadzać tę faktoryzację bo wykonujemy ją tylko raz i możemy ją wykorzystać do wyliczania wielu równań z tą samą macierzą, podczas gdy eliminację Gaussa musielibyśmy powtarzać za każdym razem.

### Algorytm Crouta

Jest to praktycznie to samo co algorytm Doolittle'a tyle że uwzględniamy tutaj element podstawowy. Przydaje się to gdy na diagonalu stoi zero. Będziemy permutować wiersze/kolumny a więc dostaniemy faktoryzację LU ale spermutowaną, musimy zapamiętać te permutacje żeby potem zastosować je do elementów wektora  $b$ .

### Algorytm Thomasa

Jest to algorytm faktoryzacji LU macierzy trójdzielnej, wraz z forward substitution i backsubstitution. Musimy zachować kształt macierzy trójdzielnej więc nie możemy użyć elementu podstawowego. Złożoność obliczeniowa rozwiązywania układu równań z macierzą trójdzielną wynosi  $O(N)$ .

### Faktoryzacja Cholesky'ego

Gdy nasz macierz jest symetryczna i dodatnio określona to istnieje faktoryzacja postaci  $A = CC^T$  gdzie  $C$  jest macierzą trójkątną dolną o elementach diagonalnych większych od zera. Znalezienie tej faktoryzacji jest +- o połowę szybsze niż znalezienie LU tej samej macierzy. Tutaj też liczy się kolejność obliczeń więc wykonujemy ją „w miejscu” bez potrzeby używania dodatkowej pamięci. Nie możemy używać elementów podstawowych ze względu na symetrię.

### Faktoryzacja LDL

Jeżeli możemy przeprowadzić faktoryzację Cholesky'ego to możemy też przeprowadzić faktoryzację LDL czyli  $A = LDL^T$  gdzie  $L$  jest macierzą trójkątną dolną z jedynkami na diagonalu a  $D$  jest macierza diagonalną o dodatnich elementach. Aby znaleźć faktoryzację LDL nie musimy pierwiastkować w przeciwieństwie do faktoryzacji Cholesky'ego.

### Macierze rzadkie, wypełnienie a faktoryzacja LU i Cholesky'ego

Ekstremalnie ważne jest aby uwzględnić fakt, że nasza macierz jest rzadka przy liczeniu na niej równań. Np. faktoryzacja LU jest dokonywana w czasie liniowym dla macierzy trójdzielnej. W przypadku faktoryzacji Cholesky'ego macierzy M-diagonalnej może pojawić się niekorzystne zjawisko zwane wypełnieniem. Jeżeli sama macierz ma zera „wewnątrz” pasma, jej czynnik Cholesky'ego nie musi ich mieć co może bardzo niekorzystnie wpłynąć na wydajność numeryczną. (Zera wewnątrz pasma znaczą

tylko, że jak macierz ma elementy niezerowe tylko na diagonalu, na pierwsze kolumnie i w pierwszym wierszu to faktoryzacja Cholesky'ego będzie macierzą pełną)

### Faktoryzacja QR

Jest to faktoryzacja  $A = QR$  gdzie  $Q$  jest macierzą ortogonalną a  $R$  trójkątną górną. Dla macierzy pełnych mamy  $O(N^3)$  czyli tak jak w LU ale współczynnik przy wyrazie wiodącym jest gorszy niż dla LU, dlatego ta metoda nie jest metoda z wyboru.  $Ax = b$ ,  $QRx = b$ ,  $Rx = Q^T b$ . Koszt obliczenia tego to  $O(N^2)$ . Faktoryzacji tej dokonujemy dwoma metodami opisanymi niżej.

### Transformacja Householdera

Jest to transformacja zerująca wszystkie składowe wektora pod konkretną. Przydaje się gdy musimy wyzerować więcej niż jeden element pod diagonalą naszej macierzy aby otrzymać faktoryzację QR. Householder przelewa całą długość wektora na jego pierwszą składową i zeruje wszystkie poniżej. Kosztuje to  $O(N)$ . Łącząc takie transformacje Householdera otrzymamy macierz  $Q^T$ .

### Obroty Givensa

Obroty Givensa służą do zerowania konkretnych składowych w wektorze. Przydają się gdy np. mamy jedną składową pod diagonalą do wyzerowania aby otrzymać faktoryzację QR. Potem łącząc macierze obrotów Givensa dostaniemy macierz  $Q$ .

### Wzór Shermana-Morrisona

Wzór ten służy do wyliczenia macierzy odwrotnej do macierzy  $A$  wtedy gdy nasza macierz  $A$  jest szczególną modyfikacją macierzy  $A_1$  a odwrotność  $A_1^{-1}$  znamy. Jawnie nigdy nie konstruujemy macierzy odwrotnej. Zapewne chcemy obliczyć jakieś  $A^{-1}b$  gdzie  $b$  jest znanym wektorem, przy założeniu że łatwo potrafimy obliczyć  $A_1^{-1}b$

### Singular Value Decomposition

Jest to metoda pozwalająca na zidentyfikowanie wektorów, dla których osobliwy układ równań ma rozwiązania. Jeżeli wyznacznik macierzy to 0 układ równań z całą pewnością nie ma jednoznacznego rozwiązania, może jednak mieć rozwiązanie (a nawet nieskończenie wiele rozwiązań).

Nadokreślone układy równań to układy równań mające więcej równań niż niewiadomych. W ogólności nie ma rozwiązań, ale można znaleźć jego rozwiązanie przybliżone za pomocą SVD

### Wykład 4 | Metody iteracyjne

Opisane wyżej metody uzyskiwania rozwiązań układów liniowych nazywamy dokładnymi (bo są dokładne gdyby nie błędy zaokrągleń) ale istnieją również iteracyjne metody uzyskiwania rozwiązań układów równań liniowych. W tych metodach rozwiązanie dokładne w teorii otrzymamy w nieskończonej liczbie kroków, ale w praktyce w skończonej liczbie kroków zbliżymy się do dokładnego rozwiązania z zadaniem przybliżeniem.

Obie poniższe metody należą do kategorii  $Mx^{(k+1)} = Nx^{(k)} + b$  gdzie  $A = M - N$ . Iteracja jest zbieżna jeśli  $\det M \neq 0$  oraz jeśli promień spektralny  $M^{-1}N < 1$

### Metoda Jacobiego

Jeżeli nieco poprzeksztalcimy układ równań to dostaniemy wzory na  $x$  w kolejnych iteracjach. Jeżeli w każdej iteracji za poprzednie wartości kolejnych  $x$  będziemy brać te z poprzedniej iteracji to otrzymamy metodę Jacobiego. Dla tej metody  $M = D$  (część diagonalna),  $N = -(L + U)$  (część pod i nad diagonalą, bez przekątnej). Metoda ta jest zbieżna gdy macierz  $A$  jest silnie diagonalnie dominująca, czyli jeśli wartości bezwzględne elementów na głównej przekątnej są większe od sumy wartości bezwzględnych pozostałych elementów w danym wierszu.

## Metoda Gaussa-Seidela

Natomiast jeśli będziemy brać za każdym razem jak najnowsze wartości  $x$  to otrzymamy zrównoleżoną metodę Jacobiego, która nazywa się metodą Gaussa-Seidela. Dla tej metody  $M = D + L$  i  $N = -U$ . Metoda ta jest zbieżna jeśli macierz  $A$  jest symetryczna i dodatnio określona.

Jeżeli macierz na której iterujemy jest rzadka to obie te metody iteracyjne będą efektywne tylko i wyłącznie wtedy, gdy w ich wzorach uwzględnimy strukturę macierzy i nie będziemy niepotrzebnie mnożyć przez zera.

## Metoda gradientów sprzężonych

Jeżeli weźmiemy funkcję  $f(x) = \frac{1}{2}x^T A x - b^T x + c$  gdzie  $A = A^T$  jest symetryczna i dodatnio określona. Przy tych założeniach ta funkcja ma dokładnie jedno minimum będące minimum globalnym. Minimum to leży w punkcie spełniającym  $\nabla f = 0$ . Jeżeli macierz  $A$  jest rzadka i duża (średnio-duża) to metoda ta jest godną uwagi metodą rozwiązywania równań  $Ax = b$ .

Conjugate Gradients (CG) – to algorytm którego istotą jest konstruowanie dwóch ciągów wektorów  $\{x_k\}$  i  $\{r_k\}$  z czego te drugie są wzajemnie prostopadłe, a zatem w arytmetyce dokładnej  $r_{N+1} = 0$  wobec czego  $x_{N+1}$  jest poszukiwanym ścisłym rozwiązaniem. Ten wariant nazywamy algebraicznym bo przy założeniu że znamy macierz  $A$  i wektor  $x_1$  możemy skonstruować ciągi  $\{r_k, p_k, x_k\}$  metodami algebraicznymi. Dla macierzy pełnej koszt tej metody wynosi  $O(N^3)$ , ale dla rzadkich jest mniejszy  $O(M \cdot N^2)$  gdzie  $M$  jest szerokością pasma.

Jeżeli współczynnik uwarunkowania macierzy  $A$  jest znacznie większy od 1 to zbieżność CG może być bardzo wolna, można temu przeciwdziałać dzięki prewarunkowaniu tej metody. Musimy wziąć sobie macierz  $C$  będącą odwracalną macierzą symetryczną rzeczywistą dodatnio określoną. Wtedy mamy  $\tilde{A} = C^{-1}AC^{-1}$  czyli  $C^{-1}AC^{-1}Cx = C^{-1}b$  czyli  $\tilde{A}\tilde{x} = \tilde{b}$ .

Niech rozkład  $QR$  macierzy  $C$  ma postać  $C = QH^T$ , gdzie  $Q$  jest macierzą ortogonalną a  $H^T$  jest macierzą trójkątną górną. Macierz  $H$  jest czynnikiem Cholesky'ego macierzy  $M = C^2 = C^T C$ . Niech rozkład Cholesky'ego macierzy  $A$  ma postać  $A = GG^T$ , przypuśćmy że  $H \cong G$  wtedy  $\tilde{A} = 1$  (macierz jednostkowa) a więc współczynnik uwarunkowania tej macierzy powinien być bliski jedności.

Jeżeli macierz  $A$  nie jest symetryczna i dodatnio określona to zakładając że jej wyznacznik jest różny od zera możemy ją „zsymetryzować” na dwa sposoby:

- CGNR:  $A^T A x = A^T b$
- CGNE:  $AA^T y = b, x = A^T y$

Do obu powyższych równań formalnie możemy używać metody gradientów sprzężonych ale trzeba pamiętać że mimo iż  $A$  jest rzadka to  $AA^T$  czy  $A^T A$  już rzadkie być nie muszą, a co więcej ich współczynnik uwarunkowania jest kwadratem współczynnika uwarunkowania macierzy wyjściowej. Alternatywnie zamiast symetryzować ta macierz możemy zmodyfikować algorytm tak żeby nie liczył dwóch ale cztery ciągi wektorów. Jednak dla wielu typów macierzy taki algorytm bywa bardzo wolno zbieżny a czasami nawet nie dostarczy rozwiązania.

## Wykład 5 | Numeryczne zagadnienie własne

Jest to problem szukania wartości własnych macierzy. Liczbę  $\lambda$  nazywamy wartością własną macierzy  $A$  jeżeli istnieje wektor  $x$  taki że  $Ax = \lambda x$ , wektor taki nazywamy wektorem własnym macierzy  $A$  do wartości własnej  $\lambda$ . Macierze rzeczywiste (niesymetryczne) mogą mieć zespolone wartości własne.

**Algorytm PageRank** – to algorytm opracowany przez twórców Google Sergeya Brina i Larry'ego Page'a, służy do obliczania rankingu stron WWW do pozycjonowania ich w przeglądarce. Uwzględnia on ilość stron linkujących do danej strony i jakąś stałą proporcjonalności. Sprowadza się do zagadnienia własnego. Wektor rankingów to wektor własny macierzy  $P$ .

**Równanie charakterystyczne macierzy  $A$**  to  $\det(A - \lambda I) = 0$  gdzie  $I$  to macierz jednostkowa. Jest to równanie wielomianowe stopnia  $N$ . Każda wartość własna jest pierwiastkiem równania charakterystycznego (ale nie koniecznie na odwrót!) Zbiór wszystkich rozwiązań równania charakterystycznego macierzy  $A$  nazywamy widmem macierzy  $A$ . Szukanie wartości własnych poprzez rozwiązywanie tego równania numerycznie jest nieefektywna i potrzebujemy innego sposobu.

### Metoda potęgowa

Jeżeli  $A$  jest macierzą symetryczną to wiemy że jest diagonalizowalna, ma wartości własne rzeczywiste. Metoda ta polega na podnoszeniu macierzy do kolejnych potęg  $k$  i dla dostatecznie dużej potęgi w równaniu otrzymamy wartość dążącą do wektora własnego do największej wartości własnej. Polega na wzięciu jakiegoś wektora startowego i w kolejnych iteracjach będziemy zbliżać się coraz to mniejszymi krokami do wektora własnego. Jeżeli weźmiemy sobie wektor startowy który jest prostopadły do uprzednio znalezionej największej wartości własnej to będziemy szukać kolejnego z kolei największego wektora własnego (innego niż pierwszy). Jeśli mamy znaleźć więcej niż jeden wektor własny to ta metoda staje się bardzo kosztowna z uwagi na konieczność reortogonalizacji. Jeśli wartości własne w macierzy są bardzo bliskie co do modułu to zbieżność będzie bardzo wolna a nawet można spodziewać się stagnacji. Metoda ta nie działa dla macierzy niesymetrycznych.

### Odwrotna metoda potęgowa

Jeżeli będziemy operować na macierzy odwrotnej to znajdując największą wartość własną macierzy odwrotnej znajdziemy najmniejszą wartość własną macierzy wyjściowej. A tak to postępujemy praktycznie tak samo jak w metodzie potęgowej tylko nie obliczamy jawnie macierzy odwrotnej!

### Transformacja podobieństwa

Dwie macierze  $A, B$  nazywamy podobnymi gdy istnieje taka nieosobliwa macierz  $S$  że  $B = S^{-1}AS$ . Macierze podobne mają te same widma.

### Ortogonalna transformacja podobieństwa

Jest to taka transformacja że  $B = O^T A O$  gdzie  $O$  jest macierzą ortogonalną.

W przypadku faktoryzacji QR widzimy że wymnożenie czynników w odwróconej kolejności stanowi ortogonalną transformację podobieństwa macierzy wyjściowej. Procedurę tę można iterować, w każdej iteracji macierz  $A_n$  jest podobna do macierzy  $A$ . Ten algorytm zachowuje symetrię, postać trójdagonalną symetryczną i postać Hessenberga.

Jeżeli macierz  $A$  jest diagonalizowalna i jej wszystkie wartości własne są rzeczywiste i parami różne od siebie to takie iterowanie jest zbieżne do macierzy trójkątnej górnej w której na głównej przekątnej stoją kolejne wartości własne.

Faktoryzacja QR ma koszt  $O(N^3)$  na każdą iterację więc jest ekstremalnie droga, ale tańsza dla macierzy rzadkich.

Jeżeli macierz nie jest symetryczna to algorytmy Givensa i Householdera z ortogonalnymi transformacjami podobieństwa dla macierzy trójdagonalnej nie prowadzą do macierzy trójdagonalnej ale do macierzy Hessenberga.



### Algorytm diagonalizacji

Diagonalizacja małych i średnich macierzy odbywa się w ten sposób, że za pomocą transformacji householdera doprowadzamy macierz do najprostszej postaci (trójdzielnej symetrycznej albo Hessenberga – zależy czy jest symetryczna). Potem taką macierz diagonalizujemy za pomocą algorytmu QR, gdyż dla tych postaci faktoryzacja QR przeprowadzana w każdym kroku algorytmu jest tania.

### Uwagi

Jeżeli mamy znać wektory własne to musimy akumulować wszystkie wykonane transformacje podobieństwa. Jeżeli mamy znać tylko wartości własne to nie musimy tego robić. Dla dobrze rozdzielonych wartości własnych algorytm QR jest szybko zbieżny, ale gdy te wartości leżą blisko siebie to zbieżność może być powolna.

### Macierze Hermitowskie

Macierz ta ma rzeczywiste wartości własne, jeżeli rozłożymy sobie ją na macierz z częścią urojoną i rzeczywistą to zauważymy, że dostaniemy macierz o podwójnej ilości wierszy i kolumn. Macierz  $H$  ma  $N$  wartości własnych ale ta nowa ma ich  $2N$ . W tym przejściu każda wartość własna się klonuje.

### Rezolwenta

Pewna liczba zespolona  $z$  nie należy do widma  $A$ . Wynika stąd, że  $\det(A - zI) \neq 0$ . Macierz  $Z = (A - zI)^{-1}$  nazywamy rezolwentą macierzy  $A$  czyli funkcją argumentu  $z$ .

### Transformacja Mobiusa

Służy do przyspieszenia poszukiwania dwóch bardzo zbliżonych wartości własnych leżących gdzieś „w środku” widma. Używa się w nim rezolwent – zbliżone wartości własne macierzy odpowiadają dobrze rozseparowanym wartościom własnym rezolwenty, wektory własne zaś są takie same.

Dobre rozseparowanie wartości własnych rezolwenty powinno sprawić, że procedura poszukiwania tych wartości własnych powinna być szybko zbieżna.

### Uogólnione zagadnienie własne

To zagadnienie  $Ax = \lambda Bx$ . Liczby  $\lambda$  nazywamy uogólnionymi wartościami własnymi, a wektory  $x$  uogólnionymi wektorami własnymi. Zagadnienia tego typu pojawiają się przy rozwiązywaniu pewnych problemów fizycznych. Jeżeli  $B$  jest symetryczna i dodatnio określona to ma faktoryzację Cholesky’ego.

W przypadku macierzy symetrycznych uogólnione wektory własne są ortogonalne względem iloczynu skalarnego generowanego przez symetryczną i dodatnio określoną macierz  $B$ .

### Wykład 6 | Interpolacja

Interpolacja to wyznaczanie wartości funkcji pomiędzy znanymi węzłami interpolacyjnymi.

Ekstrapolacja to wyznaczanie wartości funkcji w punktach leżących poza przedziałem zawierającym wszystkie węzły.

### Interpolacja odcinkami liniowa

Jest to poprowadzenie linii prostych łączących sąsiadujące punkty (węzły). Jest to nieeleganckie i w pewnych przypadkach może powodować problemy gdyż ta funkcja ma ostrza w węzłach.

### Interpolacja wielomianowa

Znając  $N$  węzłów możemy jednoznacznie wyznaczyć wielomian  $N-1$  stopnia przechodzący przez te węzły. Używając macierzy Vandermonde’a, wektora z współczynnikami  $a$  i wektora wynikowego z wartościami funkcji w danych punktach  $f$ . Możemy stworzyć układ równań którego rozwiązaniem są współczynniki wielomianu przechodzącego przez wszystkie węzły. Wyznacznik macierzy



Vandermonde'a jest różny od zera jeżeli żadne punkty  $x_1, x_2, \dots, x_n$  nie pokrywają się. Korzystając z symetrii macierzy Vandermonde'a układ taki można rozwiązać w  $O(N^2)$ .

### Wzór interpolacyjny Lagrange'a

Jest to poszukiwanie wielomianu stopnia  $N-1$  przechodzącego przez dane punkty za pomocą zadanego wzoru. W tym wzorze występuje reszta/błąd interpolacji który jednak znika tożsamościowo jeżeli  $f(x)$  (czyli poszukiwany wielomian) jest wielomianem stopnia  $N-1$ . Mówimy wtedy że ta interpolacja ma dokładność  $N-1$ . Koszt wyinterpolowania wartości w  $m$  punktach  $n$  wielomianów wynosi  $O(\max(n^2, m \cdot n))$ .

### Oscylacje Rungego

Wielomiany wysokiego stopnia są sztywne, narzucając im warunki przechodzenia przez konkretne węzły powodujemy czasami nieporządany efekt oscylacji Rungego na krańcach przedziału. Są to wahania funkcji. Występuje przy użyciu wzoru interpolacyjnego Lagrange'a. Oznacza to że interpolowanie wielomianami wysokiego stopnia jest niewskazane.

### Błąd interpolacyjny

Trudność w jego oszacowaniu polega na trudności oszacowaniu wysokich pochodnych interpolowanych funkcji.

### Współczynniki wielomianu interpolacyjnego

Współczynnik  $a_0$  moglibyśmy obliczyć obliczając wartość funkcji w zerze. Musimy wtedy wykonać  $O(N^2)$  operacji, ale następnych współczynników już tak obliczyć nie możemy bo musielibyśmy dzielić przez zero. Ale możemy użyć do tego interpolacji, konstruujemy nowy wielomian interpolacyjny i z jego pomocą obliczamy kolejne współczynniki wielomianu wyjściowego. To pokazuje że całkowity koszt wyliczenia współczynników wielomianu interpolacyjnego wynosi  $O(N^2)$ . Chyba że któryś węzeł interpolacji jest równy zero. Wtedy musielibyśmy stworzyć kolejny wielomian interpolacyjny i obliczyć jego wartość w punkcie bliskim temu węzłowi który równa się zero, wtedy zastępujemy węzeł który równa się zero tym nowym lekko przesuniętym i liczymy dalej.

### Interpolacja Hermite'a

Jeżeli poza węzłami znamy wartości pochodnej w tych punktach to mamy  $2n$  warunków na wielomian interpolacyjny. Można wtedy skonstruować interpolację wielomianową rzędu  $2n - 1$ .

### Interpolacja za pomocą funkcji sklejanych (splajny)

Unikamy tutaj efektu oscylacji Rungego. Polega to na tym że używamy funkcji sklejanych rzędu  $k$  czyli takich które są lokalnie wielomianem rzędu  $k$  i są  $(k - 1)$ -krotnie różniczkowalne w węzłach (z czego wynika, że jej pochodne rzędu  $k - 2$  i niższych są ciągłe). Najczęściej używa się splajnów rzędu 3 czyli kubicznych.

### Splajny kubiczne

Zakładając że znamy węzły i drugie pochodna wyrażenia interpolacyjnego  $\xi$  możemy skonstruować splajny kubiczne (wielomiany 3 stopnia) między węzłami interpolacyjnymi. Muszą one gładko wchodzić w siebie nawzajem. Potem każdy przedział interpolujemy za pomocą odpowiedniego splajnu. Wartości  $\xi$  otrzymamy z faktu, że żądamy iż pierwsza pochodna musi być ciągła. Po utworzeniu równania dostaniemy trójdzielny układ, który w równoodległych węzłach interpolacji jest bardzo łatwo policzyć.

### Naturalny splajn kubiczny

Jeżeli mamy  $n$  węzłów interpolacji, to mamy  $n - 2$  wewnętrznych punktów zszycia w których żądamy ciągłości pochodnej. W takim przypadku mamy układ  $n - 2$  równań z  $n$  niewiadomymi. Musimy dodać

dotatkowe warunki, domyślnie przyjmuje się że  $\xi_1 = \xi_n = 0$ . Wtedy mówimy że są to naturalne splajny kubiczne. W praktyce współczynniki  $\xi$  obliczamy tylko raz, a potem generujemy tyle splajnow kubicznych ile potrzebujemy i każdy przedział interpolujemy odpowiednim splajnem.

### Funkcje sklejane Catmulla-Roma

Jest to metoda polegająca na „dynamicznym” generowaniu splajnow pomiędzy punktami wraz z ich pojawianiem się. Służy często w grafice do generowania gładkich krzywych między punktami, gdzie kolejne punkty generowane są w trakcie postępu obliczeń więc nie znamy ich z góry i nie wiemy ile ich będzie. Używa się tutaj parametru zwanego naprężeniem który zazwyczaj ma wartość  $\frac{1}{2}$ .

### Splajny bikubiczne

Jest to interpolacja na płaszczyźnie, czyli gdy mamy funkcję dwóch zmiennych  $f(x, y)$  i tabelkę jej wartości w danych węzłach, to przeprowadzamy splajn wzdłuż każdego wiersza (po  $y$ ). Wtedy  $y$  jest stały więc tworzymy splajn jednowymiarowy kosztem  $O(N)$  ale że musimy ich wytworzyć w  $N$  wierszach mamy koszt  $O(N^2)$ . Potem obliczamy wartość każdego z powyższych splajnow w jakimś punkcie  $x^*$ . Dostaniemy  $n$  wartości funkcji w punktach  $(x^*, y_1)$ . Przez te punkty przeprowadzamy splajn w kierunku  $y$  i wyliczamy wartość tego splajnu w punkcie  $(x^*, y^*)$ . Wymaga to dodatkowych  $O(N)$  operacji więc cały koszt jest zdominowany przez  $O(N^2)$ .

### Interpolacja za pomocą funkcji wymiernych

Możemy uniknąć sztywności interpolacji wielomianowej używając do interpolowania funkcji składających się z ilorazu wielomianów. One są bardziej „giętkie”, lepiej modelują większe bogactwo zachowań. Istnieje dużo podejść do interpolacji funkcjami wymiernymi

### Algorytm Floatera i Hormanna

Jest to algorytm służący do interpolacji funkcjami wymiernymi. Sprawdzamy czy  $x$  jest blisko węzła  $x_k$ , jeśli tak to wynikiem jest stabelaryzowana wartość funkcji  $f_k$ . Jeżeli nie to obliczamy  $r(x)$  według wzoru. Wagi  $w_k$  obliczamy tylko raz na początku całej procedury. Używa się tutaj parametru interpolacji  $0 \leq d \leq n$ , zazwyczaj przyjmuje się  $d = 3$  ale czasami potrzebujemy nawet  $d = 8$ . Algorytm ten jest numerycznie mniej złożony niż splajny.

### Różniczkowanie numeryczne

Jest to obliczanie pochodnej funkcji w danym  $x$  numerycznie. Jest to bardzo podatne na błędy i trzeba go unikać. Granice we wzorze na pochodną możemy zastąpić przez:

- Iloraz różnicowy do przodu:  $f'_j = \frac{f_{j+1} - f_j}{h}$
- Wsteczny iloraz różnicowy:  $f'_j = \frac{f_j - f_{j-1}}{h}$
- Symetryczny iloraz różnicowy:  $f'_j = \frac{f_{j+1} - f_{j-1}}{2h}$

Wszystkie te przybliżenia mogą dać różne wyniki. Niesymetryczne przybliżenia mogą wprowadzić systematyczny błąd zależny od wypukłości, dlatego ostatni iloraz różnicowy jest najbezpieczniejszy pod tym względem. Im większy krok tym przybliżenie pochodnej jest gorsze.

### Różniczkowanie splajnow

Najlepszym i numerycznie tanim sposobem różniczkowania jest poprowadzenie splajnu a następnie zróżniczkowanie go w węzłach. Korzystamy przy tym z wszystkich własności splajnow, a więc z semi-analitycznych wzorów i z pewności, że pochodna jest ciągła w węzłach, co pozwala uniknąć niejednoznaczności związanej ze stosowaniem wzorów z punktu wyżej.

## Wykład 7 | Całkowanie numeryczne

Całkowanie numeryczne jest to obliczanie wartości całki numerycznie. Numerycznie wolno obliczać całki o których wiemy że istnieją. To że procedura numeryczna daje skończony (a nawet sensowny) wynik nie stanowi dowodu, że obliczana całka istnieje. Np. jeżeli będziemy chcieli scałkować całkę rozbieżną to dostaniemy jakiś wynik ale on nie jest prawidłowy.

### Kwadratury

Wzory na całkowanie przybliżone to kwadratury. Uzyskujemy je przez całkowanie wielomianów interpolacyjnych. Z interpolacji Hermite'a uzyskujemy kwadratury Gaussa, z interpolacji wielomianu Lagrange'a dostaniemy kwadratury Newtona-Cortesa. Jeżeli krańce przedziału całkowania są węzłami interpolacyjnymi to dostaniemy zamknięte kwadratury Newtona-Cortesa. W praktyce te kwadratury opiera się na interpolacji wielomianami niskiego stopnia bo łatwo się je liczy, trudno liczy się pochodne wysokich rzędów i obawiamy się oscylacji Rungego + dostaniemy większą dokładność stosując kwadratury złożone.

Najczęściej stosowane kwadratury (im większe  $n$  tym bardziej błąd rośnie, najbardziej optymalne  $n$  to 2 lub 3):

### Metoda trapezów

Kwadratura Newtona-Cortesa z  $n = 1$ . Polega na przeprowadzaniu prostej przechodzącej przez dwa węzły i obliczania pola ze wzoru na pole trapezu.

### Metoda Simpsona

Kwadratura Newtona-Cortesa z  $n = 2$ . Polega na przybliżaniu całki wielomianem drugiego stopnia. Określony jest on na trzech węzłach czyli dwóch przedziałach. Znane węzły  $x_1, x_2, x_3$  przybliża się wielomianem Lagrange'a i całkuje na przedziale  $[x_1, x_3]$ . Liczba podprzedziałów musi być parzysta.

### Metoda 3/8

Kwadratura Newtona-Cortesa z  $n = 3$ . Jest to analogiczna sytuacja do metody Simpsona, tylko że przybliżamy wielomianem trzeciego stopnia a nie drugiego. Wtedy wielomian kubiczny jest określony przez cztery punkty (trzy przedziały). Liczba podprzedziałów musi być podzielna przez 3.

### Metoda Milne'a

Kwadratura Newtona-Cortesa z  $n = 4$ . Używamy w tym przypadku pięciu węzłów, czyli cztery przedziały. Poprowadzamy wielomian czwartego stopnia i całkujemy.

Kwadratury złożone – polega na dzieleniu przedziału całkowania na podprzedziały i stosowaniu do każdego z nich kwadratur niższego rzędu zamiast stosowania kwadratur wyższego rzędu dla całego przedziału całkowania. Przedziały dzielimy w taki sposób, żeby nie obliczać ponownie tych samych wartości funkcji. Szczególnie łatwo jest to osiągnąć, gdy krańce przedziału całkowania są węzłami kwadratury, stąd bierze się popularność zamkniętych wzorów Newtona-Cortesa. Stosowanie kwadratur złożonych prowadzi do dodatkowego zmniejszenia błędu.

### Złożony wzór trapezów

Pozwala na zapamiętanie sumy wartości funkcji w poprzedniej iteracji całkowania (przed zagęszczeniem przedziałów) i stosowaniu jej w kolejnym całkowaniu. Procedurę iteracyjnego zagęszczenia przedziałów kończymy gdy kolejne znalezione przybliżenia całki różnią się od siebie zaniedbywalnie mało.

## Ekstrapolacja Richardsona

Stosując kwadratury złożone dostajemy cały ciąg kolejnych przybliżeń całki, odpowiadających kolejnym zagęszczeniom podziału. Cały ten ciąg kolejnych przybliżeń całki możemy wykorzystać przy numerycznym obliczaniu całki. Monotoniczność ciągu przybliżeń oznacza, że funkcja podcałkowa nie zmienia swojej wypukłości w przedziale całkowania a zatem jej krzywizna być może nie ulega znacznym zmianom, wobec czego założenie o stałości drugiej pochodnej w przedziale całkowania być może nie jest drastycznie złamane. Jeśli ciąg otrzymanych przybliżeń nie jest monotoniczny to stosowanie ekstrapolacji Richardsona jest wątpliwe.

## Metoda Romberga

Błąd metody trapezów zawiera wyłącznie parzyste potęgi średnicy podziału. Wykorzystując ten fakt i obliczając całkę za pomocą ciągu złożonych wzorów trapezów możemy wyjść poza ekstrapolację Richardsona uzyskując szczególnie efektywne przybliżenie całki. Tworzymy tzw. trójkąt Romberga i obliczamy kolejno jego wartości mnożąc ilość przedziałów przez 2. Jest to około 8x szybsze od metody Simpsona. (uzyskuje taką samą dokładność całki jak metoda Simpsona 8x szybciej).

## Całkowanie po przedziałach nieskończonych

Musimy uważać żeby nie obliczyć całki która jest rozbieżna. Tak to liczymy całkę od 0 do jakiegoś  $A$  i drugą od  $A$  do nieskończoności. Gdzie  $A$  jest dostatecznie dużą stałą dodatnią. Musimy to tak dobrać żeby ta druga całka do nieskończoności była do obliczenia analitycznie, tą pierwszą liczymy numerycznie. Dodajemy je do siebie i mamy wynik.

## Kwadratury adaptacyjne

Kwadratury adaptacyjne dostosowują się (krok całkowania) do funkcji którą całkują czyli do jej zmienności. Algorytm musi mieć dwa niezależne oszacowania całki po danym przedziale, ich różnica jest miarą popełnionego błędu. Analizując ten błąd i jeśli nie spełnia oczekiwań wybieramy odpowiednio przedziały i ponawiamy całkowanie dla nowego „poprawnego” przedziału. Natomiast jeśli błąd spełnia oczekiwania to całkujemy dalej wracając do „defaultowego” przedziału. Algorytm może się załamać jak przepełnimy stos przedziałów (albo osiągniemy max wysokości) to znaczy że albo całka jest rozbieżna albo zażądaliśmy nierealistycznie dużej dokładności. Musimy wtedy zwiększyć dopuszczalny błąd. Kwadratury adaptacyjne nie są najlepszą metodą obliczania całki, używa się ich do całkowania bardzo szybko oscylujących funkcji.

## Całki wymiarowe

Jeżeli wymiar całki  $n \geq 3$  to musimy ją obliczać metodami Monte Carlo które są wówczas najbardziej efektywne. Ale dla  $n = 2$  możemy stworzyć całkę iteracyjną i iterować najpierw po  $x$  a potem po  $y$  (albo na odwrót) i wtedy przynosi to dobre rezultaty.

## Kwadratury adaptacyjne w wielu wymiarach

Jeżeli mamy obszar całkowania  $D$  który jest wielokątem narzuca nam się sposób całkowania numerycznego jakim jest kwadratura adaptacyjna oparta o triangulację obszaru całkowania. Jest to pokrycie obszaru całkowania trójkątami w taki sposób że one albo się nie stykają, albo mają wspólną krawędź, albo wspólny wierzchołek. Triangulacji wstępnej można dokonać ręcznie a potem za pomocą środka ciężkości już istniejących trójkątów tworzymy 3 nowe trójkąty. Potem mając te trójkąty liczymy przybliżenie całki zgodnie ze wzorem graniastópów. Dzielimy trójkąt na trójkąty potomne i liczymy całkę po nich. Wykonujemy to dopóki stosunek powierzchni aktualnego trójkąta do powierzchni całego obszaru całkowania jest mniejszy od zadanego błędu, potem bierzemy kolejny trójkąt i analogicznie. Jak błąd jest za duży to dzielimy trójkąty i analogicznie. Robimy tak dopóki nie przerobimy wszystkich trójkątów.

## Wykład 8 | Rozwiązywanie równań algebraicznych

Rozwiązać równanie numerycznie to znalezienie jakiegoś, dowolnego rozwiązania czyli takiego  $x$  że  $f(x) = 0$ . Zakładamy że  $f$  jest ciągła i na ogół różniczkowalna odpowiednią ilość razy.

Mówimy że  $x$  jest miejscem zerowym funkcji  $f$  o krotności  $k$  gdy w tym punkcie funkcja zeruje się wraz ze swoimi pochodnymi do rzędu  $k - 1$ . Funkcja zmienia znak w otoczeniu miejsca zerowego o krotności nieparzystej i nie zmienia znaku w otoczeniu miejsca zerowego o krotności parzystej.

Metody szukania miejsc zerowych:

### Metoda bisekcji

Aby metoda zadziałała musimy mieć pewność, że funkcja zmienia znak na badanym przedziale  $[a, b]$ . Wybieramy punkt  $p$  leżący po środku tego przedziału. Jeżeli  $f(p) > 0$  to powtarzamy operację dla przedziału  $[a, p]$ , w przeciwnym przypadku dla przedziału  $[p, b]$ . Iterujemy tak aż otrzymamy zadaną dokładność. Metoda ta nie działa dla miejsc zerowych o parzystej krotności!

### Metoda reguła fałsi

W tej metodzie również funkcja musi zmienić znak na badanym przedziale, więc ta metoda nie zadziała dla miejsc zerowych o parzystych krotnościach. W tej metodzie gdy badamy przedział  $[a, b]$  to prowadzimy sobie prostą z punktu  $(a, f(a))$  do punktu  $(b, f(b))$  i badamy punkt w którym ta prosta przetnie się naszą badaną funkcją  $f$ . Jeżeli wartość funkcji w tym punkcie jest większa od 0 to bierzemy przedział na lewo i powtarzamy operację, w przeciwnym wypadku bierzemy przedział na prawo i powtarzamy operację. Iterujemy tak aż do otrzymania zadanej dokładności.

### Metoda siecznych

Zaczynamy od dwóch punktów  $a$  i  $b$  takich że  $f(a) \neq f(b)$ . Prowadzimy sieczną przez te punkty (bez względu na ich znak) i jako punkt  $p$  bierzemy miejsce zerowe tej siecznej. W kolejnych iteracjach bierzemy zawsze dwa ostatnio wyliczone punkty, bez względu na to czy ta funkcja zmienia znak. Ta metoda czasami nie jest zbieżna do miejsca zerowego.

### Interpolacja odwrotna

Polega na odwróceniu funkcji i stworzeniu wielomianu interpolacyjnego funkcji odwrotnej. Wartość funkcji odwrotnej w zerze oznacza punkt, w którym funkcja ma miejsce zerowe. Używając wielomianu interpolacyjnego liczymy wartość tego wielomianu czyli przybliżenia funkcji odwrotnej w zerze. Stosujemy to dla niewielkiej liczby węzłów, wynik interpolacji odwrotnej może robić za punkt startowy innych bardziej dokładnych metod.

### Kryterium stopu

Metody wyżej są metodami iteracyjnymi więc muszą mieć jakieś kryterium stopu, kiedy muszą się zatrzymać. Kryterium naiwne to porównywanie wartości funkcji w danym punkcie do zera i badanie jak duża jest różnica. Możemy też również jako kryterium stopu obrać przedział badany i zatrzymać iterowanie gdy jest on dostatecznie mały. Poza tym możemy założyć maksymalną liczbę iteracji dla metody która nie musi być zbieżna, uchroni nas to przed nieskończonymi pętlami.

### Metoda Newtona

Inaczej metoda stycznych. Startujemy z pewnego punktu. Kolejne punkty wyznaczamy ze wzoru  $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$ . Nazwa wzięła się stąd, że po prostu tworzymy styczną do wykresy funkcji i przechodzącą przez nasz punkt i badamy miejsce w którym przetnie się z osią OX, nazwijmy ten punkt  $p$ . Potem startując z punktu  $(p, f(p))$  tworzymy nową styczną itd. aż dostaniemy dostatecznie dobre miejsce zerowe. Metoda ta jednak może być rozbieżna i czasami prowadzi do cykli. Dodatkowo jeżeli

miejsca zerowe są wielokrotne to metoda Newtona jest zbieżna liniowo, dla miejsc zerowych jednokrotnych jest zbieżna kwadratowo. Trzeba ustalić maksymalną liczbę iteracji ze względu na możliwość nie zbiegnięcia i/lub cykli.

Basen atrakcji jest to zbiór punktów o tej własności, że metoda Newtona zastartowana z takiego punktu prowadzi do wskazanego miejsca zerowego.

Granice basenów atrakcji w metodzie Newtona na płaszczyźnie zespolonej to bardzo często są fraktale.

#### Tłumiona metoda Newtona

Aby uciec z wielocyklu możemy zamiast metody Newtona wykonać tłumioną metodę Newtona czyli pomnożyć ten iloraz z wzoru wyżej przez jakąś stałą z przedziału  $(0, 1]$ . Aby uciec z wielocyklu robimy takie 2-3 kroki tą metoda.

#### Metody wykorzystujące drugą pochodną

Metoda Newtona wykorzystuje rozwinięcie Taylora do pierwszego rzędu, możemy to uogólnić na rozwinięcie do drugiego rzędu. Dojdziemy znowu do wzoru iteracyjnego w którym znak w mianowniku wybieramy tak, aby moduł mianownika był większy. W odróżnieniu od standardowej metody Newtona, ta może prowadzić do zespolonych iteratów także dla rzeczywistych wartości początkowych.

#### Metoda Halleya

Jest to inna metoda w której stosujemy metode Newtona do równania:  $g(x) = \frac{f(x)}{\sqrt{|f'(x)|}} = 0$ . Wtedy każdy pierwiastek  $f$  który nie jest miejscem zerowym pochodnej, jest pierwiastkiem  $g$ . Każdy pierwiastek  $g$  jest pierwiastkiem  $f$ .

#### Układy równań algebraicznych

Rozwiązanie takich układów jest trudne gdyż geometrycznie oznacza znalezienie punktu przecięcia krzywych. O tych funkcjach na ogół nic nie wiemy, zmiana jednej nie wpływa na zmianę innej itd.

#### Wielowymiarowa metoda Newtona

Rozwijając funkcję  $g$  w szereg Taylora do pierwszego rzędu otrzymamy równanie z jacobianem funkcji  $g$ . Dostaniemy iterację  $x_{k+1} = x_k - J^{-1}(x_k)g(x_k)$ . Nie należy konstruować jawnej odwrotności jacobianu. Zapis  $z = J^{-1}g$  rozumiemy jako  $Jz = g$ . Tutaj w każdym kroku musimy rozwiązywać jacobian czyli kolejny układ równań liniowych co czyni tę metodę kosztowną. Często dla przyspieszenia obliczeń jacobian zmieniamy nie co krok ale co kilka kroków.

#### Rozwiązywanie równań nieliniowych a minimalizacja

Rozwiązywanie równań algebraicznych jest „trudne” a minimalizacja „łatwa” ale funkcja  $g$  może mieć wiele minimów lokalnych więc nie mamy gwarancji, że globalne minimum istnieje. Nie jest to więc dobry pomysł rozwiązywać równania algebraiczne jak minimalizacje.

#### Metoda globalnie zbieżna

Połączenie idei minimalizacji funkcji i metody Newtona. Kierunek kroku Newtona jest lokalnym kierunkiem spadku funkcji. Jednak przesunięcie o pełną długość kroku Newtona nie musi prowadzić do spadku funkcji. Ta metoda jest zawsze zbieżna do jakiegoś minimum funkcji ale niekoniecznie do jej minimum globalnego. Jeżeli znajdziemy minimum lokalne z pożądaną tolerancją to należy rozpocząć z innym warunkiem początkowym. Jeżeli kilka różnych warunków początkowych nie daje rezultatu to się poddajemy. Szansa na znalezienie numerycznego rozwiązania układu równań jest tym większa im lepszy jest warunek początkowy, więc musimy go bardzo dobrze wybrać.



Do tych metod potrzebujemy analitycznych wzorów na pochodne funkcji bo liczenie ich w sposób numeryczny na ogół nie ma sensu.

#### Wielowymiarowa metoda siecznych – metoda Broydena

Czasami analityczne wzory na pochodne są nieznane i samo obliczenie jakobianu jest numerycznie zbyt kosztowne. W takich przypadkach używa się tej metody. Tutaj pochodną zastępujemy ilorazem różnicowym. Metoda ta wymaga inicjalizacji przez podanie  $B_1$  i  $x_1$ . To drugie nie jest niczym dziwnym, co do pierwszego jeśli to możliwe można przyjąć  $B_1 = J(x_1)$

#### Wykład 9 | Miejsca zerowe wielomianów

Rozwiązywanie równań wielomianowych jest jednym z nielicznych przypadków, w których chcemy poznać wszystkie pierwiastki równania nieliniowego.

Obliczanie wartości wielomianu w sposób naiwny jest bardzo kosztowne, znacznie bardziej efektywne jest wyliczanie kolejnych potęg za pomocą mnożenia i obliczanie wielomianu począwszy od wyrazu wolnego.

**Algorytm Hornera** – polega na obliczaniu wartości wielomianu w danym punkcie, posiada  $N$  mnożeń i  $N$  dodawań. Przekształcamy wielomian wyciągając przed nawias kolejne  $x$ .

Zazwyczaj nie znamy dokładnych współczynników wielomianu których pierwiastków poszukujemy. Najczęściej są one obarczone jakimś błędem bo pochodzą z poprzednich obliczeń.

#### Przykład Wilkinsona

Jeżeli weźmiemy wielomian  $W(x) = (x + 1)(x + 2) \dots (x + 20)$  to zauważymy że jego miejscami zerowymi są liczby całkowite ujemne  $-1, -2, \dots, -20$ . Jeżeli zaburzymy jeden współczynnik w nieznacznym sposób to zaburzenie miejsca zerowego jest siedem rzędów wielkości większe od zaburzenia pojedynczego współczynnika. W rzeczywistości miejsca zerowe tak zaburzonego wielomianu stają się nawet zespolone. Zagadnienie znajdowania miejsc zerowych wielomianów może być źle uwarunkowane.

W przypadku wielokrotnych miejsc zerowych zaburzenie współczynników o bardzo małe odchylenie może powodować przeniesienie się miejsc zerowych z osi rzeczywistej na płaszczyznę zespoloną.

Jeśli otrzymamy grupę leżących blisko siebie „numerycznych miejsc zerowych” czasami trudno jest rozstrzygnąć czy są one naprawdę różne, czy też na skutek skończonej dokładności z jaką znamy współczynniki reprezentują one rozszczerzone miejsce wielokrotne.

Aby efektywnie szukać tych miejsc zerowych musimy mieć dobrą metodę numeryczną do ich szukania (czyli metody Laguerre’a) i musimy obniżać (deflacja) stopień wielomianu po każdorazowym znalezieniu jego miejsca zerowego przez dzielenie wielomianu przez  $(x - x_0)$  i wygładzeniu znalezionych miejsc zerowych przy pomocy pierwotnego niewydzielonego wielomianu.

#### Metoda Laguerre’a

Jest to metoda iteracyjna do poszukiwania miejsc zerowych wielomianu. Znak w mianowniku wybieramy tak aby był większy. W wypadku ogólnym metoda ta jest zbieżna sześciennie do wszystkich pojedynczych miejsc zerowych (rzeczywistych i zespolonych), jest zbieżna liniowo do wielokrotnych miejsc zerowych i przypadki braku zbieżności są rzadkie. Jeżeli tej metodzie grozi stagnacja to można wykonać 1 lub 2 kroki metodą Newtona i potem powrócić do Laguerre’a.

Metoda ta jest metodą z wyboru przy poszukiwaniu rzeczywistych i zespolonych, pojedynczych i wielokrotnych miejsc zerowych wielomianów. Wymaga ona obliczania drugiej pochodnej ale w



wypadku wielomianów jest to bardzo proste. Nawet dla rzeczywistych punktów początkowych może prowadzić do zespolonych miejsc zerowych.

Metoda Laguerre'a jest podobna do metody opartej o rozwinięcie w szereg Taylora do drugiego rzędu ale jest od niej lepsza bo uwzględnia stopień wielomianu.

Aby uniknąć sytuacji gdy metoda zbiega nam do tego samego miejsca zerowego (mimo innego punktu początkowego) możemy obniżyć stopień wielomianu. W tym przypadku do wygładzenia wielomianu będziemy szukać nowych współczynników i dostaniemy układ równań z macierzą posiadającą jedynki na diagonalu i  $-x_0$  pod diagonalą. Bardzo łatwo to wyliczyć metodą forward substitution.

Szukanie miejsc zerowych wielomianu, wygładzanie, faktoryzacje i obniżanie stopnia wykonujemy aż nasz wielomian nie będzie stopnia drugiego. Wtedy już mamy analityczne wzory na miejsce zerowe i z nich korzystamy.

Jeżeli wyjściowy wielomian ma współczynniki rzeczywiste to wiemy że miejsca zerowe albo są rzeczywiste albo tworzą zespolone pary sprzężone. Temu jak znajdziemy jedno miejsce zerowe będące liczbą zespoloną to automatycznie możemy wyznaczyć drugie przez sprzężenie tej liczby.

Czasami aby znaleźć pierwiastki wielomianu szukamy wartości własnych macierzy stowarzyszonej (companion matrix). Równanie charakterystyczne tej macierzy różni się od ogólnego wielomianu tylko tym, że współczynnik przy najwyższej potęgzie wynosi 1. Każdy wielomian stopnia  $N$  można sprowadzić do tej postaci dzieląc przez współczynnik przy najwyższej potęgzie. Nie zmienia to pierwiastków wielomianu.

## Wykład 10 | Minimalizacja funkcji jednej zmiennej

Minimalizacja to poszukiwanie minimów lokalnych (nie globalnych) funkcji ciągłych których argument także zmienia się w sposób ciągły. Jeżeli chcemy znaleźć maksimum funkcji  $g(x)$  to szukamy minimum funkcji  $f(x) = -g(x)$ , ponieważ minimum tej funkcji jest maksimum funkcji wyjściowej.

W tym przypadku do stwierdzenia że funkcja ma minimum w jakimś punkcie nie wystarczy obliczenie wartości funkcji w tym punkcie, musimy porównywać je z uprzednio policzonymi wartościami w innych punktach. Procedurę poszukiwania minimum kończymy wtedy gdy zmiana wartości funkcji staje się porównywalna z zadaną tolerancją, w szczególności, z dokładnością, z jaką prowadzimy obliczenia.

Minimum jesteśmy w stanie zlokalizować z dokładnością co najwyżej równą pierwiastkowi z numerycznej precyzji obliczeń.

Wstępna lokalizacja minimum mówimy że punkty  $(a, b, c)$  ograniczają (lokalizują) minimum funkcji  $f(x)$  jeśli

$$a < b < c: f(a) > f(b), f(c) > f(b)$$

Wyznaczamy wartość funkcji w dwóch punktach, wyznaczają one kierunek spadku funkcji. Wyznaczamy trzeci punkt idąc w kierunku spadku o taką samą odległość, jaka dzieliła punkty początkowe. Jeżeli warunek wyżej nie jest spełniony to podwajamy krok, zawsze biorąc pod uwagę dwa ostatnio obliczone punkty. Trzeba ograniczyć maksymalną dopuszczalną liczbę kroków (lub wielkość kroków) żeby zabezpieczyć się przed nieskończoną pętlą.

Mając trójkę liczb  $(a, b, c)$  wybieramy pewien punkt  $d \neq b$  taki że  $a < d < c$  jako nowy punkt wewnętrzny. Potem w zależności od  $f(d)$  wybieramy albo lewy albo prawy przedział i powtarzamy iteracje. Jeżeli tylko funkcja jest ciągła na przedziale  $[a, c]$  i ma w nim minimum to ta strategia zbierze się do tego minimum. Lub jeśli ma ich więcej to zbierze do jednego z nich.

### Metoda złotego podziału

W tej metodzie punktu  $d$  szukamy w większym z dwóch podprzedziałów  $[a, b]$  i  $[b, c]$ . Poszczególne przedziały powinny spełniać złotą proporcję: długość większego podprzedziału ma się do długości całego przedziału tak, jak długość mniejszego do długości większego. Metoda ta jest zbieżna liniowo.

### Metoda paraboliczna i metoda Brenta

Metodą lepszą od złotego podziału jest interpolacja paraboliczna (bo efektywniej wykorzystuje wartości funkcji). Przez trzy punkty przeprowadzamy parabolę a jako punkt  $d$  bierzemy jej minimum. Dalej zaś postępujemy analogicznie. W pobliżu minimum funkcja powinna z dobrym przybliżeniem zgadzać się ze swoim rozwinięciem w szereg Taylora do drugiego rzędu. Dzięki temu ta metoda może być zbieżna szybciej niż liniowo, ale mogą się zdarzyć kłopoty, zwłaszcza jeśli początkowo funkcja nie przypomina paraboli. W związku z tym stosujemy metodę Brenta:

Obliczamy  $d$  i jeśli  $d$  leży wewnątrz przedziału  $a < d < c$  oraz rozmiar nowego przedziału wyznaczonego przez metodę opisaną w wstępnym lokalizowaniu minimum jest mniejszy niż połowa przedziału w przedostatniej iteracji to akceptujemy to  $d$ . W innym przypadku nie akceptujemy  $d$  i dokonujemy bisekcji przedziału  $[a, c]$  i jako  $d$  bierzemy jego środek a dalej postępujemy analogicznie.

Metoda Brenta jest uznawana za podstawową metodę poszukiwania minimów funkcji jednowymiarowych.

### Wykorzystanie pochodnych

Jeżeli obliczanie pochodnej funkcji  $f$  jest łatwe i jest ona różniczkowalna to możemy posłużyć się pochodną do przyspieszenia zbieżności. Ale w przeciwieństwie do funkcji wielu zmiennych, wykorzystanie pochodnej nie powoduje znacznego przyspieszenia a niekiedy zwłaszcza daleko od minimum metoda ta może zachowywać się gorzej niż metoda złotego podziału. Najprostszą metodą jest dokonanie interpolacji liniowej pochodnych w skrajnych punktach przedziału. Jako punkt  $d$  bierzemy miejsce zerowe funkcji interpolującej pochodną.

### Wykorzystanie interpolacji Hermite'a

Można też użyć bardziej złożonej metody wykorzystującej pochodne: korzystając z interpolacji Hermite'a. Konstruujemy wielomian trzeciego stopnia zgadzający się z badaną funkcją i jej pochodną w skrajnych punktach przedziału, po czym jako punkt  $d$  bierzemy minimum tego wielomianu leżące w przedziale  $[a, c]$  o ile takowe istnieje.

Metoda Brenta i metody wykorzystujące pochodne przyspieszają zbieżność jeżeli funkcja daje się dobrze przybliżyć za pomocą rozwinięcia Taylora do rzędu drugiego.

## Wykład 11 | Minimalizacja funkcji wielu zmiennych

To podstawowe zastosowanie metod numerycznych. Stosujemy dwie różne strategie:

- Gdy jesteśmy daleko od poszukiwanego minimum szybko podążamy w jego kierunku, nie starając się jednak znaleźć dokładnego położenia minimum.
- Gdy jesteśmy blisko poszukiwanego minimum i zależy nam na jego dokładnym zlokalizowaniu staramy się o taką dokładną lokalizację, wiedząc że może ona być kosztowna nawet w problemach kilkowymiarowych.

Podstawową zasadą jest to że idziemy zawsze w dół, czyli w stronę malejących wartości funkcji. Znalezienie minimum oznacza rozwiązanie równania  $\nabla f = 0$

### Metoda najszybszego spadku (steepest descent, gradient descent)

Gdy jesteśmy daleko od minimum to staramy się o to aby w kolejnych krokach naszej procedury wartości funkcji malały. Poruszamy się w kierunku przeciwnym do gradientu funkcji. Ta metoda jest nieprecyzyjna.

### Stochastic Gradient Descent

Czasami czas obliczania sumy gradientów cząstkowych może być bardzo znaczny. Jeżeli założymy że poszczególne elementy tej sumy nie odbiegają zbyt od średniej to możemy średni, kosztowny w wyliczaniu gradient zastąpić losowo wybranym gradientem cząstkowym. Wtedy otrzymujemy właśnie tą metodę. Jeżeli nasz losowo wybrany gradient cząstkowy jest gorszy od poprzedniego to losujemy ponownie aż otrzymamy lepszy (Stochastic Gradient Descent z ograniczeniem). Iteracje kończymy albo po określonej ilości kroków albo gdy dostaniemy zadaną dokładność. Metoda ta działa bardzo dobrze dla bardzo dużych zbiorów uczących. Przyjęcie warunku iż wartość funkcji  $Q$  musi maleć po wykonaniu kroku na ogół pogarsza wyniki. Od czasu do czasu można/trzeba wykonać krok „w złą stronę”.

### Zastosowanie metody Newtona

Można spróbować rozwiązać równanie  $\nabla f = 0$  przy pomocy wielowymiarowej metody Newtona. Pojawi się nam hessjan czyli macierz drugich pochodnych cząstkowych funkcji  $f$ . Dostaniemy wzór iteracyjny, jednakże jest pewien problem z tą metodą. Dla dużych  $N$  zastosowanie tej metody jest monstrualnie kosztowne. Nawet dla  $N$  rzędu kilku-kilkunastu wielokrotne wyliczanie pochodnych cząstkowych a następnie rozwiązywanie układu równań może być drogie i uciążliwe. Jeżeli minimalizowana funkcja nie jest kwadratowa (a zazwyczaj nie jest) to krok Newtonowski może być zbyt długi i może przestrzelić nad poszukiwanym minimum. Warunkiem koniecznym aby ta iteracja prowadziła do zmniejszenia wartości funkcji jest symetria i dodatnia określoność wszystkich napotkanych po drodze hessianów.

### Dodatnia określoność hessjanu

Możemy założyć że jeśli minimalizowana funkcja jest dostatecznie gładka to hessjan jest w każdym punkcie symetryczny. Niestety nie można tego założyć odnośnie do dodatniej określoności. W minimum hessjan jest dodatnio określony. Na podstawie ciągłości drugich pochodnych wnioskujemy że musi istnieć pewne otoczenie minimum, w którym hessjan jest dodatnio określony. Jednak daleko od minimum hessjan dodatnio określony być nie musi. Możemy więc traktować dodatnią określoność hessjanu jako matematyczny warunek tego, że jesteśmy blisko minimum.

### Metoda Levenberga-Marquardta

Metodę tą stosujemy gdy wymiarowość problemu nie jest bardzo duża (100 co najwyżej), obliczanie drugich pochodnych jest możliwe i niezbyt uciążliwe i gdy koszt obliczania drugich pochodnych nie jest wielki ze względu na strukturę problemu. Metoda ta wywodzi się z metody Newtona.

Daleko od minimum nie minimalizujemy tylko podążamy w kierunku malejących wartości funkcji.

### Strategia precyzyjnej minimalizacji wielowymiarowej

Na podstawie dodatniej określoności hessjanu albo wiedząc że krajobraz daleko od minimum ma „strukturę lejka” określamy czy jesteśmy dostatecznie blisko minimum. Jeżeli tak to można a niekiedy wręcz trzeba zastosować metody, które pozwolą na bardziej precyzyjną i możliwie szybką lokalizację minimum. Strategia oszukiwania (lokanego) minimum funkcji w postaci ciągu minimalizacji jednowymiarowych polega na szukaniu po kolei minimów funkcji jednowymiarowych aż znajdziemy minimum ogólne. Możemy albo minimalizować po współrzędnych (kolejnymi kierunkami poszukiwań są kierunki równoległe do kolejnych osi układu współrzędnych) albo metodą najszybszego spadku (w której kierunek poszukiwań pokrywa się z minus gradientem minimalizowanej funkcji w danym

punkcie). Jednak jeżeli jesteśmy bardzo blisko minimum to strategie te spowodują bardzo liczne małe kroki.

#### Funkcja Rosenbrocka

$f(x, y) = (1 - x)^2 + 100(y - x^2)^2$  jest to tzw. funkcja Rosenbrocka będąca dobrym przykładem funkcji trudnej do zminimalizowania.

#### Metoda gradientów sprzężonych

Pozwala rozwiązać te układy których macierze są symetryczne i dodatnio określone. Forma kwadratowa  $\frac{1}{2}x^T Ax - b^T x + f(0)$  ma jedno minimum dla  $Ax = b$ , które jest w punkcie  $\nabla f = 0$  i jest równoważne rozwiązaniu układu. Zmiana gradientu musi być prostopadła do starego kierunku poszukiwać i tak być musi dla wszystkich poprzednich kierunków. W tej metodzie te kroki, które wymagałyby znajomości hessjanu w minimum zastępujemy minimalizacją kierunkową, natomiast te kroki, które nie wymagają znajomości hessjanu wykonujemy tak samo, jak w algebraicznej metodzie gradientów sprzężonych. Formalnie daje to to samo co algebraiczna metoda gradientów sprzężonych.

#### Metoda zmiennej metryki

Polega na stosowaniu ciągu dodatnio określonych przybliżeń hessjanu zamiast wyliczania za każdym razem hessjanu dokładnego. Metoda zmiennej metryki ma się do metody gradientów sprzężonych tak jak metoda Broydena ma się do wielowymiarowej metody Newtona.

#### Metoda Powella

Metoda Powella polega na konstrukcji kierunków, które z czasem po wielu iteracjach stają się sprzężone. Stosujemy ją gdy obliczanie pochodnych jest kosztowne, niemożliwe lub funkcja jest nieróżniczkowalna.

#### Wykład 12 | Uwagi o minimalizacji globalnej

Czasami musimy znaleźć więcej minimów lub minimum globalne, wtedy poprzednie algorytmy są nieprzydatne. Algorytmy globalnej optymalizacji bardzo często zawierają element losowości bo nie wiemy gdzie znajduje się minimum globalne. Strategie zachłanne nie przydają się do szukania minimów globalnych funkcji (bardzo) wielu zmiennych.

#### Algorytm Monte Carlo

Używamy N-wymiarowej gaussowskiej zmiennej losowej do szukania minimum globalnych. Nie mamy naturalnego kryterium stopu, wykonujemy tyle kroków ile ustalimy. Dobrze jest kilkakrotnie wylosować punkt startowy i odpalić algorytm dla różnych punktów a nie wykonywać dużo kroków dla jednego punktu. Dla zbyt dużej ilości wymiarów nie ma gwarancji, że wylosowane punkty same z siebie znajdą się w basenie atrakcji wszystkich interesujących minimów. Domyślnie przyjmujemy prawdopodobieństwo Gauss-Boltzmann. Parametry określają jak łatwo algorytmowi jest „iść pod górę”. Dobieramy je metodą prób i błędów i wcześniejszymi doświadczeniami.

Metody tej można używać do problemów dyskretnych! Niestety czasami ta metoda może – zwłaszcza w przypadku problemów bardzo wielo-wymiarowych – wpaść w pułapkę minimum lokalnego.

#### Algorytmy genetyczne

Są inspirowane podstawową zasadą ewolucyjną: survival of the fittest (najlepiej dostosowani przeżywają). W kontekście algorytmów genetycznych często zamiast o minimalizacji mówi się o maksymalizacji funkcji dopasowania, zwanej także funkcją celu. W tym algorytmie mamy do czynienia z całą populacją osobników. Jeden osobnik reprezentuje stan optymalizowanego układu. Każdy osobnik reprezentowany jest przez chromosom. W algorytmie tym obliczamy wartość funkcji

dopasowania dla wszystkich członków populacji, następnie wybieramy tych członków którzy są najbardziej dopasowani i tworzymy z dwojga losowych najbardziej dopasowanych osobników nich nowe pokolenie. Przekazujemy im chromosomy rodziców ale czasami, rzadko może zdarzyć się mutacja, przestawienie czegoś w chromosomie. Iteracje kończymy gdy wykorzystamy z góry założoną liczbę epok lub gdy wszyscy członkowie populacji (chromosomy) znajdują się w niewielkim otoczeniu jakiegoś stanu, lub gdy funkcja dopasowania przestaje się zmieniać na przestrzeni kilku epok.

### Paralelizacja

Polega na użyciu wątków do generowania potomków a następnie scalaniu wątków i obliczaniu funkcji dopasowania i prawdopodobieństwa wyboru dla członków populacji potomnej. Potem przechodzimy do kolejnej epoki. Zapamiętujemy chromosom, który w ciągu wszystkich epok dał najlepszą wartość funkcji dopasowania.

### Particle Swam Optimization

Jest to kolejna klasa algorytmów służąca do minimalizacji funkcji zmiennych zmieniających się w sposób ciągły. Jest ona także motywowana biologicznie, w tym wypadku obserwacją zachowań stad ptaków czy ławic ryb. Mamy tutaj do czynienia z rojami osobników. Osobniki w pewne sposoby oddziałują ze sobą i mają swojakom „inteligencję”. Osobniki w roju mogą porozumiewać się ze sobą wskazując na przykład basen atrakcji czy znalezione minima itp. Mogą kierować się w stronę minimum globalnego i mają pewną prędkość. Optymalnie jest stworzyć kilka rojów w różnych miejscach na płaszczyźnie w celu szukania miejsc zerowych niezależnych od siebie.

### Wykład 13 | Aproksymacja i zagadnienie najmniejszych kwadratów

Aproksymacja punktowa to szukanie funkcji należącej do znanej kategorii która będzie przebiegać możliwie najbliżej  $N$  punktów. Funkcja jest znana co do swojego kształtu (wielomian pewnego stopnia, kombinacja funkcji trygonometrycznych itp.) a tylko nieznane są jej parametry.

Aproksymacja ciągła to mając ustaloną funkcję  $g(x)$  której sposób obliczania jest trudny staramy się skonstruować inną funkcję, która będzie w pewnym sensie bliska wyjściowej a jednocześnie obliczeniowo prostsza.

Skrajnym przypadkiem aproksymacji punktowej jest interpolacja. Nie chcemy takiego efektu bo mimo iż funkcja przechodzi wtedy przez wszystkie punkty to jest „trudną” funkcją.

### Liniowe zagadnienie najmniejszych kwadratów

Przyjmujemy że wartość teoretyczna jest kombinacją liniową pewnych znanych funkcji:

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

Zespół wszystkich wartości teoretycznych możemy zatem przedstawić jako  $\tilde{y} = Ap$ . Problemem numerycznym, który chcemy rozwiązać jest znalezienie najlepszego wektora parametrów  $p$ .

### Błędy pomiarowe

Różnica pomiędzy wartością zmierzoną a wartością teoretyczną jest spowodowana błędem pomiarowym.

### Metoda najmniejszych kwadratów

Jeżeli błędy pomiarowe pochodzą z rozkładu Gaussa o macierzy kowariancji  $G$ , estymator największej wiarygodności odpowiada minimum formy kwadratowej  $Q = \frac{1}{2} \xi^T G^{-1} \xi$ . Obecność odwrotności macierzy kowariancji oznacza że pomiary obciążone większym błędem dają mniejszy wkład do  $Q$ .

## Forma kwadratowa estymatorów

W liniowym zagadnieniu najmniejszych kwadratów minimalizowana funkcja jest formą kwadratową w parametrach. Dzięki temu wiemy, że minimum istnieje i jest jednoznaczne. Liniowość oznacza tutaj, że funkcja „teoretyczna” zależy liniowo od parametrów, nie od argumentu.

## Minimum formy kwadratowej

Aby znaleźć estymator, należy znaleźć taki wektor  $p$ , że forma kwadratowa przybiera najmniejszą możliwą wartość. Można to zrobić albo bezpośrednio, metodą zmiennej metryki lub gradientów sprzężonych albo formalnie rozwiązując równanie  $\nabla Q = 0$  gdzie różniczkujemy po składowych wektora  $p$ . Otrzymujemy  $A^T G^{-1} A p = A^T G^{-1} y$ . Tego równania nie da się uprościć pozbywając się członu  $A^T G^{-1}$  bo nie jest to macierz kwadratowa – dla niej nie da się zdefiniować odwrotności.

## Nadokreślony układ równań

Zamiast minimalizować formę kwadratową moglibyśmy zażądać aby równanie  $y_i = \tilde{y}_i$  było ściśle spełnione dla wszystkich punktów pomiarowych. Wówczas chcemy rozwiązać układ liniowy z  $s$  niewiadomymi i  $n$  równań gdzie  $n > s$ . Dzięki SVD możemy uzyskać przybliżone rozwiązanie tego układu. W praktyce zdarza się często że przybliżone rozwiązanie uzyskane tą metodą jest tym samym rozwiązaniem które otrzymalibyśmy minimalizując formę kwadratową lub rozwiązując układ równań.

## Pomiary nieskorelowane

Na ogół zakłada się, że pomiary są niezależne a ich wyniki nieskorelowane. Wówczas elementy pozadiagonalne macierzy  $G$  znikają. Minimalizowana forma kwadratowa się upraszcza.

## Kryterium Akaike

Czasami nie wiadomo ile funkcji bazowych należy uwzględnić w dopasowaniu. W szczególności jeśli do danych doświadczalnych dopasowujemy wielomian, niekiedy – jeśli nie mamy dobrego modelu teoretycznego – nie wiemy jaki stopień wielomianu wybrać. Jest jasne, że im wyższy stopień wielomianu tym dopasowanie będzie „lepsze” ale zawsze staramy się dobrać model o jak najmniejszej liczbie parametrów. Hirotugu Akaike zaproponował kryterium które nagradza za jak najlepsze dopasowanie ale karze za zbyt wiele parametrów. Należy zminimalizować wielkość  $AIC = \ln Q + \frac{2s}{N}$  gdzie  $Q$  jest wartością minimalizowanej formy kwadratowej w minimum zwaną błędem rezydualnym,  $s$  jest liczbą parametrów,  $N$  liczbą punktów do których dopasowujemy.

## Nieliniowe zagadnienie najmniejszych kwadratów

Przypuśćmy że dopasowywana do danych pomiarowych zależność teoretyczna zależy od parametrów w sposób nieliniowy.  $\tilde{y}_i = f(x_i; p)$  gdzie  $p$  jest wektorem parametrów. Zakładamy że  $f(\cdot; p)$  jest znaną funkcją, a tylko jej parametry są nieznane. Na przykład do danych doświadczalnych dopasowujemy funkcję Gaussa.

## Pseudolinearyzacja

Czasami do znalezienia minimum  $Q$  stosuje się metodę pseudolinearyzacji. Przypuśćmy, że  $p$  jest aktualnym przybliżeniem poszukiwanej wartości parametrów. Stawiamy hipotezę, iż „prawdziwe” wartości parametrów są małą poprawką w stosunku do  $p$ . Rozwijamy to w szereg Taylora do pierwszego rzędu i po paru przestawieniach i podstawieniach dojdziemy do funkcji, która jest formą kwadratową w poprawkach  $\delta p$ . Po znalezieniu znanymi metodami minimum tej funkcji odpowiadających (jedynemu) minimum  $Q$  podstawiamy  $p_{n+1} = p_n + \delta p_{min}$  i powtarzamy całą procedurę. To dobrze działa w wypadku nieliniowej metody najmniejszych kwadratów, choć nie należy jej polecać jako ogólnej metody minimalizacji. Po znalezieniu ostatecznych wartości minimalizujących funkcję za pomocą pseudolinearyzacji wokół tego punktu znajdujemy macierz kowariancji estymatorów, będącą charakterystyką liniową.

## Wykład 14 | Przybliżenia Pade

Jest to zagadnienie obliczania aproksymacji ciągłej. Przypuśćmy, że znamy wartości pewnej funkcji  $f$  i jej pochodnych do rzędu  $N$  w zerze (lub innym punkcie, można je łatwo przesunąć do zera). Na tej podstawie możemy skonstruować przybliżenie funkcji  $f$  w pewnym przedziale zawierającym zero, tak aby to przybliżenie zgadzało się z funkcją i jej  $N$  pochodnymi w zerze.

Najprostszym sposobem jest skonstruowanie rozwinięcia Maclaurina do rzędu  $N$ . Otrzymujemy w ten sposób wielomian, który co prawda spełnia wymagania w otoczeniu zera ale przybliżenie wielomianowe zazwyczaj szybko załamuje się już w niewielkiej odległości od zera. Lepsze jest przybliżenie wymierne.

Czasami jednak te przybliżenia nie istnieją. Np. funkcja  $\cosh(x)$  jest parzysta, jej rozwinięcie Maclaurina też jest parzyste, więc przybliżenia nie spełniające warunku parzystości nie mogą istnieć.

Lepsze przybliżenie otrzymujemy używając ilorazu wielomianów Czebyszewa. Są one lepsze od przybliżeń Pade gdyż błąd tych pierwszych zachowuje się bardziej regularnie w całym przedziale.

made by Karosek <3