Optymalizacja ciągła

Imię i Nazwisko: .... **Numer indeksu:** ....

Test zaliczeniowy, grupa: $\boxed{\text{B}}$

1. `[2pkt]` Przy każdym zdaniu zaznacz, czy jest prawdziwe (`P`), czy fałszywe (`F`):
   (a) W minimum globalnym funkcji różniczkowalnej (znajdującym się wewnątrz dziedziny) zeruje się gradient.
   (b) Liniowy rząd zbieżności oznacza, że błąd optymalizacji maleje w tempie wykładniczym.
   (c) Kwadratowy rząd zbieżności oznacza spadek błędu optymalizacji odwrotnie proporcjonalny do kwadratu liczby iteracji
   (d) Dwukrotnie różniczkowalne funkcje wypukłe mają w każdym punkcie dodatnio półokreślony hesjan
   (e) Funkcja wypukła osiąga minimum globalne w dokładnie jednym punkcie dziedziny.
   (f) Punkt stacjonarny z hesjanem o wartościach własnych mających różne znaki wyklucza minimum lokalne

2. `[2pkt]` Przy każdym zdaniu dotyczącym metod jednowymiarowych zaznacz, czy jest prawdziwe (`P`), czy fałszywe (`F`):
   (a) Znalezienie minimum globalnego dowolnej funkcji jednowymiarowej jest możliwe z dowolną dokładnością w logarytmicznej liczbie kroków.
   (b) Metoda złotego podziału wymaga mniej odwołań do funkcji w każdej iteracji niż metoda przeszukiwania dychotomicznego.
   (c) Metoda bisekcji zawsze znajduje minimum dowolnej funkcji wielomianowej w liczbie kroków równej stopniowi tego wielomianu.
   (d) Do znalezienia minimum funkcji w danym przedziale wymagane jest by była ona różniczkowalna w tym przedziale.
   (e) Metoda bisekcji do działania wymaga by funkcja była różniczkowalna na badanym przedziale.

3. `[3pkt]` Używając operacji zachowujących wypukłość, pokaż, że funkcja $f\colon \mathbb{R}^3 \to \mathbb{R}$:
$$f(x,y,z) = e^{\max\{x+z,y\}} - \ln(x+y),$$
   jest wypukła.
   Odp:

4. `[4pkt]` Rozważ minimalizację funkcji dwóch zmiennych:
$$f(x,y) = \frac{1}{2}(x+y)^2 - 8\sqrt{x} - 4\sqrt{y}.$$
   Rozpoczynając od punktu $(x_0, y_0) = (1,1)$ wykonaj jeden krok:
   - metodą spadku wzdłuż gradientu z ustaloną długością kroku $\alpha = \frac{1}{2}$,
   - metodą Newtona-Raphsona.

   Wyznacz pochodne cząstkowe $f(x,y)$:

   Stąd $\nabla f(x_0, y_0) =$

   Wyznacz drugie pochodne cząstkowe $f(x,y)$:

   Stąd $\nabla^2 f(x_0, y_0) =$

   $\left(\nabla^2 f(x_0, y_0)\right)^{-1} =$

   Rozwiązanie po pierwszym kroku algorytmu spadku wzdłuż gradientu (napisz wzór ogólny i wartość liczbową):

   $\boldsymbol{x}_1 = (x_1, y_1) =$

   Pierwszy krok algorytmu Newtona-Raphsona (napisz wzór ogólny i wartość liczbową):

   $\boldsymbol{x}_1 = (x_1, y_1) =$

5. [2pkt] Przy każdym zdaniu dotyczącym metody gradientów sprzężonych (MGS) zaznacz, czy jest prawdziwe (P), czy fałszywe (F):

   (a) Wektory $\begin{bmatrix}1\\1\end{bmatrix}$ i $\begin{bmatrix}0\\1\end{bmatrix}$ są sprzeżone względem $\boldsymbol{A} = \begin{bmatrix}2 & -1\\-1 & 1\end{bmatrix}$.

   (b) Wektory $\begin{bmatrix}2\\-1\end{bmatrix}$ i $\begin{bmatrix}1\\2\end{bmatrix}$ są sprzeżone względem $\boldsymbol{A} = \begin{bmatrix}3 & 2\\2 & 3\end{bmatrix}$.

   (c) MGS znajduje dokładne rozwiązanie optymalne funkcji kwadratowej z dodatnio określoną macierzą w n krokach (gdzie n to liczba zmiennych).

   (d) Kierunek sprzężony dobrany w danej iteracji przez MGS może nie być kierunkiem poprawy.

   (e) Kierunki sprzężone są do siebie ortogonalne.

   (f) Metody Fletchera-Reevesa i Polaka-Ribière'a są identyczne dla funkcji kwadratowej.

6. [2pkt] Przy każdym zdaniu dotyczącym omówionych metod spadkowych zaznacz, czy jest prawdziwe (P), czy fałszywe (F):

   (a) Metoda najszybszego spadku wzdłuż gradientu gwarantuje spadek funkcji celu w każdej iteracji (o ile nie zeruje się gradient).

   (b) Metoda Levenberga-Marquardta ze zmienną długością kroku gwarantuje spadek funkcji celu w każdej iteracji (o ile nie zeruje się gadient).

   (c) Duży wskaźnik uwarunkowania macierzy wypukłej funkcji kwadratowej możne oznaczać wolną zbieżność algorytmu Cauchy'ego.

   (d) Metoda Newtona-Raphsona zmniejsza wartość funkcji celu w każdym kroku.

   (e) Brak kierunku spadkowego w punkcie wewnątrz dziedziny funkcji różniczkowalnej implikuje punkt stacjonarny.

   (f) Metoda Newtona-Raphsona i spadku wzdłuż gradientu działają identycznie dla funkcji kwadratowej.

7. [2pkt] Przy każdym zdaniu dotyczącym metody stochastycznego spadku wzdłuż gradientu (SGD) zaznacz, czy jest prawdziwe (P), czy fałszywe (F):

   (a) Algorytm SGD zwykle nie porusza się w kierunku ujemnego gradientu funkcji celu.

   (b) Algorytm SGD optymalizujący funkcję ściśle wypukłą, ze stałą długością kroku i nieskończenie dużą liczbą iteracji, znajdzie minimum globalne funkcji.

   (c) Algorytm SGD zbiega wolniej (w funkcji liczby iteracji) niż klasyczny algorytm spadku wzdłuż gradientu.

   (d) Jest duża szansa, że algorytm SGD zainicjalizowany w minimum funkcji celu w swojej pierwszej iteracji pogorszy funkcję celu.

   (e) Algorytm SGD jest metodą spadkową i w każdej iteracji zmniejsza wartość funkcji celu (o ile nie jesteśmy w optimum).

8. [3pkt] Udowodnij, że jeśli kierunek w punkcie $\boldsymbol{x}_k$ zostaje wybrany jako $\boldsymbol{v}_k = -\boldsymbol{A}\,\nabla f(\boldsymbol{x}_k)$, gdzie $\boldsymbol{A}$ jest *dowolną* macierzą dodatnią określoną, to $\boldsymbol{v}_k$ jest kierunkiem poprawy dla funkcji f. Pokaż, że implikuje to, że metoda Newtona-Raphsona jest metodą spadkową jeśli hesjan funkcji f jest dodatnio określony w każdym punkcie dziedziny.

Klucz rozwiązań

Optymalizacja ciągła, Test zaliczeniowy, grupa: $\boxed{\text{B}}$

1. `[2pkt]` Przy każdym zdaniu zaznacz, czy jest prawdziwe (`P`), czy fałszywe (`F`):

$\boxed{\text{(a)}}$ W minimum globalnym funkcji różniczkowalnej (znajdującym się wewnątrz dziedziny) zeruje się gradient.

$\boxed{\text{(b)}}$ Liniowy rząd zbieżności oznacza, że błąd optymalizacji maleje w tempie wykładniczym.

(c) Kwadratowy rząd zbieżności oznacza spadek błędu optymalizacji odwrotnie proporcjonalny do kwadratu liczby iteracji

$\boxed{\text{(d)}}$ Dwukrotnie różniczkowalne funkcje wypukłe mają w każdym punkcie dodatnio półokreślony hesjan

(e) Funkcja wypukła osiąga minimum globalne w dokładnie jednym punkcie dziedziny.

$\boxed{\text{(f)}}$ Punkt stacjonarny z hesjanem o wartościach własnych mających różne znaki wyklucza minimum lokalne

2. `[2pkt]` Przy każdym zdaniu dotyczącym metod jednowymiarowych zaznacz, czy jest prawdziwe (`P`), czy fałszywe (`F`):

(a) Znalezienie minimum globalnego dowolnej funkcji jednowymiarowej jest możliwe z dowolną dokładnością w logarytmicznej liczbie kroków.

$\boxed{\text{(b)}}$ Metoda złotego podziału wymaga mniej odwołań do funkcji w każdej iteracji niż metoda przeszukiwania dychotomicznego.

(c) Metoda bisekcji zawsze znajduje minimum dowolnej funkcji wielomianowej w liczbie kroków równej stopniowi tego wielomianu.

(d) Do znalezienia minimum funkcji w danym przedziale wymagane jest by była ona różniczkowalna w tym przedziale.

$\boxed{\text{(e)}}$ Metoda bisekcji do działania wymaga by funkcja była różniczkowalna na badanym przedziale.

3. `[3pkt]` Używając operacji zachowujących wypukłość, pokaż, że funkcja $f\colon \mathbb{R}^3 \to \mathbb{R}$:

$$f(x,y,z) = e^{\max\{x+z,y\}} - \ln(x+y),$$

jest wypukła.

Odp:

Funkcja $\max\{x+z, y\}$ jest wypukła jako maksimum funkcji wypukłych (liniowych). Dalej, $e^{\max\{x+z,y\}}$ jest wypukła jako złożenie wypukłej niemalejącej z wypukły. Również $-\ln(x+y)$ jest wypukła złożenie funkcji wypukłej z liniową, co oznacza, że $f(x,y,z)$ jest wypukła jako suma dwóch funkcji wypukłych.

4. `[4pkt]` Rozważ minimalizację funkcji dwóch zmiennych:

$$f(x,y) = \frac{1}{2}(x+y)^2 - 8\sqrt{x} - 4\sqrt{y}.$$

Rozpoczynając od punktu $(x_0, y_0) = (1,1)$ wykonaj jeden krok:

- metodą spadku wzdłuż gradientu z ustaloną długością kroku $\alpha = \frac{1}{2}$,
- metodą Newtona-Raphsona.

Wyznacz pochodne cząstkowe $f(x,y)$:

$\frac{\partial f}{\partial x} = x + y - 4/\sqrt{x}$, $\quad \frac{\partial f}{\partial y} = x + y - 2/\sqrt{y}$

$\nabla f(x_0, y_0) = (-2, 0)$

Wyznacz drugie pochodne cząstkowe $f(x,y)$:

$\frac{\partial^2 f}{\partial x^2} = 1 + 2x^{-3/2}$, $\frac{\partial^2 f}{\partial x \partial y} = 1$, $\frac{\partial^2 f}{\partial y^2} = 1 + y^{-3/2}$

$\nabla^2 f(x_0,y_0) = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$

$\left(\nabla^2 f(x_0,y_0)\right)^{-1} = \frac{1}{5}\begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix}$

Rozwiązanie po pierwszym kroku algorytmu spadku wzdłuż gradientu (napisz wzór ogólny i wartość liczbową):

$\boldsymbol{x}_1 = (x_1, y_1) = (1,1) - \frac{1}{2} \cdot (-2, 0) = (2, 1)$

Pierwszy krok algorytmu Newtona-Raphsona (napisz wzór ogólny i wartość liczbową):

$\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \frac{1}{5}\begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix}\begin{bmatrix} -2 \\ 0 \end{bmatrix} = \begin{bmatrix} 9/5 \\ 3/5 \end{bmatrix}$

5. `[2pkt]` Przy każdym zdaniu dotyczącym metody gradientów sprzężonych (MGS) zaznacz, czy jest prawdziwe (`P`), czy fałszywe (`F`):

$\boxed{\text{(a)}}$ Wektory $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ i $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ są sprzeżone względem $\boldsymbol{A} = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$.

(b) Wektory $\begin{bmatrix} 2 \\ -1 \end{bmatrix}$ i $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ są sprzeżone względem $\boldsymbol{A} = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$.

$\boxed{\text{(c)}}$ MGS znajduje dokładne rozwiązanie optymalne funkcji kwadratowej z dodatnio określoną macierzą w n krokach (gdzie n to liczba zmiennych).

(d) Kierunek sprzężony dobrany w danej iteracji przez MGS może nie być kierunkiem poprawy.

(e) Kierunki sprzężone są do siebie ortogonalne.

$\boxed{\text{(f)}}$ Metody Fletchera-Reevesa i Polaka-Ribière'a są identyczne dla funkcji kwadratowej.

6. [2pkt] Przy każdym zdaniu dotyczącym omówionych metod spadkowych zaznacz, czy jest prawdziwe (P), czy fałszywe (F):

   $\boxed{\text{(a)}}$ Metoda najszybszego spadku wzdłuż gradientu gwarantuje spadek funkcji celu w każdej iteracji (o ile nie zeruje się gradient).

   $\boxed{\text{(b)}}$ Metoda Levenberga-Marquardta ze zmienną długością kroku gwarantuje spadek funkcji celu w każdej iteracji (o ile nie zeruje się gadient).

   $\boxed{\text{(c)}}$ Duży wskaźnik uwarunkowania macierzy wypukłej funkcji kwadratowej możne oznaczać wolną zbieżność algorytmu Cauchy'ego.

   (d) Metoda Newtona-Raphsona zmniejsza wartość funkcji celu w każdym kroku.

   $\boxed{\text{(e)}}$ Brak kierunku spadkowego w punkcie wewnątrz dziedziny funkcji różniczkowalnej implikuje punkt stacjonarny.

   (f) Metoda Newtona-Raphsona i spadku wzdłuż gradientu działają identycznie dla funkcji kwadratowej.

7. [2pkt] Przy każdym zdaniu dotyczącym metody stochastycznego spadku wzdłuż gradientu (SGD) zaznacz, czy jest prawdziwe (P), czy fałszywe (F):

   $\boxed{\text{(a)}}$ Algorytm SGD zwykle nie porusza się w kierunku ujemnego gradientu funkcji celu.

   (b) Algorytm SGD optymalizujący funkcję ściśle wypukłą, ze stałą długością kroku i nieskończenie dużą liczbą iteracji, znajdzie minimum globalne funkcji.

   $\boxed{\text{(c)}}$ Algorytm SGD zbiega wolniej (w funkcji liczby iteracji) niż klasyczny algorytm spadku wzdłuż gradientu.

   $\boxed{\text{(d)}}$ Jest duża szansa, że algorytm SGD zainicjalizowany w minimum funkcji celu w swojej pierwszej iteracji pogorszy funkcję celu.

   (e) Algorytm SGD jest metodą spadkową i w każdej iteracji zmniejsza wartość funkcji celu (o ile nie jesteśmy w optimum).

8. [3pkt] Udowodnij, że jeśli kierunek w punkcie $\boldsymbol{x}_k$ zostaje wybrany jako $\boldsymbol{v}_k = -\boldsymbol{A}\ \nabla f(\boldsymbol{x}_k)$, gdzie $\boldsymbol{A}$ jest *dowolną* macierzą dodatnią określoną, to $\boldsymbol{v}_k$ jest kierunkiem poprawy dla funkcji f. Pokaż, że implikuje to, że metoda Newtona-Raphsona jest metodą spadkową jeśli hesjan funkcji f jest dodatnio określony w każdym punkcie dziedziny.

   Dowód: Aby pokazać, że $\boldsymbol{v}_k$ jest kierunkiem poprawy, trzeba udowodnić, że tworzy on kąt rozwarty z gradientem, tzn. $\nabla f(\boldsymbol{x}_k)^\top \boldsymbol{v}_k < 0$. Mamy:

   $$\nabla f(\boldsymbol{x}_k)^\top \boldsymbol{v}_k = -\nabla f(\boldsymbol{x}_k)^\top \boldsymbol{A} \nabla f(\boldsymbol{x}_k) < 0,$$

   gdzie ostatnia nierówność wynika z dodatniej określoności macierzy $\boldsymbol{A}$. W metodzie Newtona-Raphsona mamy $\boldsymbol{A} = \left(\nabla^2 f(\boldsymbol{x}_k)\right)^{-1}$, która jest dodatnio określona, ponieważ $\nabla^2 f(\boldsymbol{x}_k)$ jest dodatnio określony z założenia.