Optymalizacja ciągła

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

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

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

   (a) W minimum globalnym różniczkowalnej funkcji jednej zmiennej na przedziale $[a, b]$ zeruje się pochodna.

   (b) Liniowy rząd zbieżności oznacza, że odległość od optimum maleje liniowo z liczbą iteracji.

   (c) Kwadratowy rząd zbieżności oznacza, że błąd optymalizacji maleje w tempie podwójnie wykładniczym.

   (d) Jeśli $f(\boldsymbol{x}) \geqslant f(\boldsymbol{y}) + \nabla f(\boldsymbol{y})^\top(\boldsymbol{x} - \boldsymbol{y})$ dla dowolnych $\boldsymbol{x}, \boldsymbol{y}$ z dziedziny, to f jest wypukła.

   (e) Funkcja wypukła może osiągać minimum globalne w więcej niż jednym punkcie dziedziny

   (f) Punkt stacjonarny z dodatnio określonym hesjanem oznacza minimum lokalne.

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

   (a) Jeśli pochodna jednowymiarowej funkcji jednomodalnej równa jest zero w danym punkcie to znajduje się tam minimum.

   (b) Metoda złotego podziału wymaga średnio mniej odwołań do funkcji by znaleźć rozwiązanie od metody przeszukiwania dychotomicznego.

   (c) Dla funkcji kwadratowej przeszukiwanie dychotomiczne znajduje optymalne rozwiązanie w jednym kroku.

   (d) Metoda bisekcji jest metodą bezgradientową.

   (e) Metoda przeszukiwania dychotomicznego może nie znaleźć minimum jeśli funkcja nie jest jednomodalna.

   (f) Metoda przeszukiwania dychotomicznego wymaga aby funkcja była różniczkowalna w badanym przedziale.

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

   $$f(x, y, z) = e^{(x-y)^2} - 2x + z^2,$$

   jest wypukła.

   Odp:

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

   $$f(x, y) = e^{x+y} - 2x + y^2$$

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

   - metodą spadku wzdłuż gradientu z długością kroku $\alpha = 1$,
   - 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 $\begin{bmatrix}0\\1\end{bmatrix}$ i $\begin{bmatrix}1\\0\end{bmatrix}$ są sprzeżone względem $\boldsymbol{A} = \begin{bmatrix}2 & 1\\1 & 1\end{bmatrix}$.

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

   (c) Punkt minimum dowolnej funkcji różniczkowalnej jest kombinacją liniową wektorów sprzężonych.

   (d) MGS zastosowany do dowolnej funkcji wypukłej kończy działanie po n krokach (gdzie n to liczba zmiennych).

   (e) MGS jest metodą spadkową.

   (f) Kierunek sprzężony $\boldsymbol{v}_k$ wybrany w k-tej iteracji MGS jest prostopadły do gradientu w tej iteracji, $\boldsymbol{g}_k = \nabla f(\boldsymbol{x}_k)$.

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

   (a) Metoda najszybszego spadku wzdłuż gradientu gwarantuje znalezienie minimum globalnego funkcji różniczkowalnej.

   (b) Wskaźnik uwarunkowania macierzy wypukłej funkcji kwadratowej równy jedności gwarantuje znalezienie minimum w jednym kroku metody Cauchy'ego.

   (c) Metoda Levenberga-Marquardta i spadku wzdłuż gradientu działają identycznie dla funkcji kwadratowej.

   (d) Metoda Newtona-Raphsona zainicjowana dostatecznie blisko minimum (z dodatnio określonym hesjanem) zbiega z kwadratowym rzędem zbieżności.

   (e) Niezerowy gradient wewnątrz dziedziny funkcji oznacza istnienie w tym punkcie kierunku spadkowego.

   (f) Metoda spadku wzdłuż gradientu z regułą Armijo zastosowana do funkcji wypukłej gwarantuje spadek funkcji celu (o ile nie zeruje się gradient).

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

   (a) Jedna iteracja algorytmu SGD zainicjalizowanego w minimum funkcji nie będzie miała żadnego efektu (nie zmieni aktualnego rozwiązania).

   (b) Kierunek zmian w danej iteracji algorytmu SGD jest średnio równy ujemnemu gradientowi funkcji celu.

   (c) Algorytm SGD można zastosować do optymalizacji regresji liniowej z niekwadratowym błędem o następujące postaci: $f(\boldsymbol{w}) = \sum_{i=1}^{N} (y_i - \boldsymbol{x}_i^\top \boldsymbol{w})^4$.

   (d) Zbieżność algorytmu SGD jest taka sama jak klasycznego algorytmu spadku wzdłuż gradientu (tj. liniowa).

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

8. `[3pkt]` Pokaż, że jeśli $\boldsymbol{v}_k$ jest *dowolnym* kierunkiem poprawy w k-tej iteracji dla różniczkowalnej funkcji f, a nowy punkt $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k$ zostaje wybrany poprzez optymalizację długości kroku (tzn. $\alpha_k = \operatorname{argmin}_\alpha f(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k)$, załóż, że takie α_k istnieje), to gradient w nowym punkcie jest prostopadły do $\boldsymbol{v}_k$, tzn. $\nabla f(\boldsymbol{x}_{k+1})^\top \boldsymbol{v}_k = 0$. W szczególności, pokaż, że wynika z tego twierdzenie o "zygzakowaniu" dla metody Cauchy'ego.

Klucz rozwiązań

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

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

   (a) W minimum globalnym różniczkowalnej funkcji jednej zmiennej na przedziale $[a, b]$ zeruje się pochodna.

   (b) Liniowy rząd zbieżności oznacza, że odległość od optimum maleje liniowo z liczbą iteracji.

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

   $\boxed{\text{(d)}}$ Jeśli $f(\boldsymbol{x}) \geqslant f(\boldsymbol{y}) + \nabla f(\boldsymbol{y})^\top(\boldsymbol{x} - \boldsymbol{y})$ dla dowolnych $\boldsymbol{x}, \boldsymbol{y}$ z dziedziny, to f jest wypukła.

   $\boxed{\text{(e)}}$ Funkcja wypukła może osiągać minimum globalne w więcej niż jednym punkcie dziedziny

   $\boxed{\text{(f)}}$ Punkt stacjonarny z dodatnio określonym hesjanem oznacza minimum lokalne.

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

   (a) Jeśli pochodna jednowymiarowej funkcji jednomodalnej równa jest zero w danym punkcie to znajduje się tam minimum.

   $\boxed{\text{(b)}}$ Metoda złotego podziału wymaga średnio mniej odwołań do funkcji by znaleźć rozwiązanie od metody przeszukiwania dychotomicznego.

   (c) Dla funkcji kwadratowej przeszukiwanie dychotomiczne znajduje optymalne rozwiązanie w jednym kroku.

   (d) Metoda bisekcji jest metodą bezgradientową.

   $\boxed{\text{(e)}}$ Metoda przeszukiwania dychotomicznego może nie znaleźć minimum jeśli funkcja nie jest jednomodalna.

   (f) Metoda przeszukiwania dychotomicznego wymaga aby funkcja była różniczkowalna w badanym przedziale.

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

   $$f(x, y, z) = e^{(x-y)^2} - 2x + z^2,$$

   jest wypukła.

   Odp:

   Funkcja $x - y$ jest liniowa, stąd funkcja $(x-y)^2$ jest wypukła jako złożenie liniowej i wypukłej. Dalej, $e^{(x-y)^2}$ jest wypukła jako złożenie wypukłej niemalejącej z wypukły. Na koniec $f(x, y, z)$ jest wypukła jako suma trzech funkcji wypukłych.

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

   $$f(x, y) = e^{x+y} - 2x + y^2$$

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

   - metodą spadku wzdłuż gradientu z długością kroku $\alpha = 1$,
   - metodą Newtona-Raphsona.

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

   $\frac{\partial f}{\partial x} = e^{x+y} - 2, \quad \frac{\partial f}{\partial y} = e^{x+y} + 2y$

   $\nabla f(x_0, y_0) = (-1, -1)$

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

   $\frac{\partial^2 f}{\partial x^2} = e^{x+y}$, $\frac{\partial^2 f}{\partial x \partial y} = e^{x+y}$, $\frac{\partial^2 f}{\partial y^2} = e^{x+y} + 2$

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

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

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

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

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

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

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

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

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

   $\boxed{\text{(c)}}$ Punkt minimum dowolnej funkcji różniczkowalnej jest kombinacją liniową wektorów sprzężonych.

   (d) MGS zastosowany do dowolnej funkcji wypukłej kończy działanie po n krokach (gdzie n to liczba zmiennych).

   $\boxed{\text{(e)}}$ MGS jest metodą spadkową.

   (f) Kierunek sprzężony $\boldsymbol{v}_k$ wybrany w k-tej iteracji MGS jest prostopadły do gradientu w tej iteracji, $\boldsymbol{g}_k = \nabla f(\boldsymbol{x}_k)$.

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

   (a) Metoda najszybszego spadku wzdłuż gradientu gwarantuje znalezienie minimum globalnego funkcji różniczkowalnej.

   $\boxed{\text{(b)}}$ Wskaźnik uwarunkowania macierzy wypukłej funkcji kwadratowej równy jedności gwarantuje znalezienie minimum w jednym kroku metody Cauchy'ego.

   (c) Metoda Levenberga-Marquardta i spadku wzdłuż gradientu działają identycznie dla funkcji kwadratowej.

   $\boxed{\text{(d)}}$ Metoda Newtona-Raphsona zainicjowana dostatecznie blisko minimum (z dodatnio określonym hesjanem) zbiega z kwadratowym rzędem zbieżności.

   $\boxed{\text{(e)}}$ Niezerowy gradient wewnątrz dziedziny funkcji oznacza istnienie w tym punkcie kierunku spadkowego.

   $\boxed{\text{(f)}}$ Metoda spadku wzdłuż gradientu z regułą Armijo zastosowana do funkcji wypukłej gwarantuje spadek funkcji celu (o ile nie zeruje się gradient).

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

   (a) Jedna iteracja algorytmu SGD zainicjalizowanego w minimum funkcji nie będzie miała żadnego efektu (nie zmieni aktualnego rozwiązania).

   $\boxed{\text{(b)}}$ Kierunek zmian w danej iteracji algorytmu SGD jest średnio równy ujemnemu gradientowi funkcji celu.

   $\boxed{\text{(c)}}$ Algorytm SGD można zastosować do optymalizacji regresji liniowej z niekwadratowym błędem o następujące postaci: $f(\boldsymbol{w}) = \sum_{i=1}^{N}(y_i - \boldsymbol{x}_i^\top \boldsymbol{w})^4$.

   (d) Zbieżność algorytmu SGD jest taka sama jak klasycznego algorytmu spadku wzdłuż gradientu (tj. liniowa).

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

8. `[3pkt]` Pokaż, że jeśli $\boldsymbol{v}_k$ jest *dowolnym* kierunkiem poprawy w k-tej iteracji dla różniczkowalnej funkcji f, a nowy punkt $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k$ zostaje wybrany poprzez optymalizację długości kroku (tzn. $\alpha_k = \operatorname{argmin}_\alpha f(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k)$, załóż, że takie α_k istnieje), to gradient w nowym punkcie jest prostopadły do $\boldsymbol{v}_k$, tzn. $\nabla f(\boldsymbol{x}_{k+1})^\top \boldsymbol{v}_k = 0$. W szczególności, pokaż, że wynika z tego twierdzenie o "zygzakowaniu" dla metody Cauchy'ego.

Dowód: Niech $g(\alpha) = f(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k)$. Z definicji α_k mamy $g'(\alpha_k) = 0$ (pochodna zeruje się gdzieś ponieważ $\boldsymbol{v}_k$ jest kierunkiem poprawy, a punkt minimum α_k z założenia istnieje), czyli:

$$\begin{aligned} g'(\alpha_k) &= \left.\frac{\mathrm{d}f(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k)}{\mathrm{d}\alpha}\right|_{\alpha=\alpha_k} \\ &= \left.\left(\nabla f(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k)^\top \boldsymbol{v}_k\right)\right|_{\alpha=\alpha_k} \\ &= \nabla f(\boldsymbol{x}_{k+1})^\top \boldsymbol{v}_k = 0 \end{aligned}$$

W metodzie Cauchy'ego przyjmujemy jako $\boldsymbol{v}_k$ kierunek $-\nabla f(\boldsymbol{x}_k)$, który jest kierunkiem poprawy, ponieważ $\boldsymbol{v}_k^\top \nabla f(\boldsymbol{x}_k) = -\nabla f(\boldsymbol{x}_k)^\top \nabla f(\boldsymbol{x}_k) = -\|\nabla f(\boldsymbol{x}_k)\|^2 < 0$.