131890-S Metody Optymalizacji **Letni 2023/2024**

Zajęcia 2: Optymalizacja analityczna z ograniczeniami

Daniel Kaszyński

2.1 Własności gradientu funkcji

Pojęciem już wprowadzonym jest pojęciu gradientu funkcji – wektora pochodnych cząstkowych. Gradient funkcji będzie często wykorzystywany w ramach wykładu, stąd warto znać jego dwie, istotne z perspektywy optymalizacji, własności. Dla przypomnienia:

Definicja 1: Gradient funkcji

Niech $f : \mathbb{D} \subset \mathbb{R}^n \to \mathbb{R}$, $x \in \mathbb{D}$. Gradientem funkcji f, jest funkcja $\nabla_f(x) : \mathbb{R}^n \to \mathbb{R}^n$ w punkcie x nazywamy funkcję:

$$\nabla_f(x) = \left[\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x), \ldots, \frac{\partial f}{\partial x_n}(x)\right]$$

Pamiętajmy, że:

Związek między Pochodną Kierunkową a Gradientem? Jeżeli istnieje gradient funkcji $\nabla_f(\mathbf{x})$ w punkcie $\mathbf{x}$ (co oznacza, że f jest różniczkowalna w $\mathbf{x}$)

$$\nabla_f(\mathbf{x}) = \left[\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right]$$

to pochodna kierunkowa funkcji f w kierunku wektora $\mathbf{h}$ jest równa iloczynowi skalarnemu gradientu funkcji f i wektora $\mathbf{h}$

Twierdzenie 1: Gradient jest kierunkiem najszybszego wzrostu wartości funkcji.

Dowód. Niech $|h| = 1$, tj. będzie znormalizowanym wektorem. Wtedy stopa wzrostu wartości funkcji f w punkcie x w kierunku h jest dana przez pochodną kierunkową $\frac{df}{dh}(x)$. Wyznaczmy w takim razie kierunek h, który maksymalizuje stopę wzrostu wartości funkcji f, tj. kierunek maksymalizujący pochodną kierunkową:

$$\frac{df}{dh}(x) = \nabla_f(x)h = |\nabla_f(x)||h|cos(\nabla_f(x), h) = |\nabla_f(x)|cos(\nabla_f(x), h)$$

dla $|\nabla_f(x)| \geqslant 0$ oraz $cos(\nabla_f(x), h) \in [-1, 1]$ stopa wzrostu wartości funkcji f jest największa, gdy $cos(\nabla_f(x), h) = 1$, co implikuje, że h wskazuje ten sam kierunek co $\nabla_f(x)$. Oznacza to, że $h = \frac{\nabla_f(x)}{|\nabla_f(x)|}$. □

Twierdzenie 2: Gradient jest ortogonalny względem warstwicy funkcji.

Dowód. Niech $f : \mathbb{R}^n \to \mathbb{R}$, $x^* = (x_1^*, \dots, x_n^*)$ oraz $\nabla_f(x^*) \neq 0$. Niech $r : \mathbb{R} \to \mathbb{R}^n$ taki, że $r(t_0) = x^*$. Wartość funkcji jest stała dla wszystkich punktów z zadanej warstwicy (zgodnie z definicją warstwicy) oraz: $\forall_{t\in\mathbb{R}} f(r(t)) = c$. Wtedy $\frac{d}{dt}(f(r(t)) = \nabla_f(r(t))\frac{dr}{dt}(t) = 0$. W szczególności $\nabla_f(r(t_0))\frac{dr}{dt}(t_0) = 0$. Ponieważ $\frac{dr}{dt}(t_0)$ jest przestrzenią styczną do warstwicy funkcji f w x^*, oznacza to wtedy, że $\nabla_f(x^*)$ jest prostopadła do warstwicy funkcji. □

2.1.1 Warunki Pierwszego Rzędu w przypadku ograniczeń równościowych

Twierdzenie 3: Twierdzenie Lagrange'a, Mnożniki Lagrange'a.
Niech $f : \mathbb{D} \subset \mathbb{R}^n \to \mathbb{R}$, $f \in C^1$. Jeżeli funkcja f posiada w x ekstremum związane $h(x) = 0$, gdzie $h : \mathbb{R}^n \to \mathbb{R}^m$, $h \in C^1$, w punkcie x, to

$$\nabla_f(x) + \lambda^T \mathbf{D}h(x) = \mathbf{0}$$

Dla $h : \mathbb{R}^n \to \mathbb{R}$.

$$\nabla_f(x) + \lambda\nabla_h(x) = \mathbf{0}$$

Przykład 1. [Niech $f(x) = x_1^2 + x_2^2$, oraz $[x_1, x_2] : h(x) = x_1^2 + 2x_2^2 = 0$. Znajdź ekstrema funkcji $f(x)$ p.w. $h(x) = 1$. Z Warunków Pierwszego Rzędu (FOC) uzyskujemy:

$$\begin{cases} 2x_1 + \lambda 2x_1 = 0 \Rightarrow x_1(1+\lambda) = 0 \\ 2x_2 + \lambda 4x_2 = 0 \Rightarrow x_2(1+2\lambda) = 0 \end{cases}$$

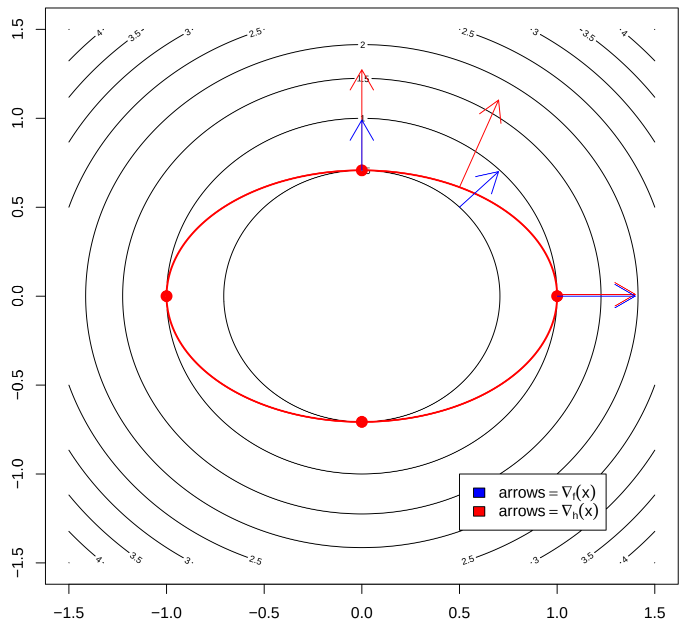
Co daje nam 4 rozwiązania:

1. $[x_1, x_2] = \left[0, \frac{1}{\sqrt{2}}\right]$, $\lambda = -\frac{1}{2}$

2. $[x_1, x_2] = \left[0, -\frac{1}{\sqrt{2}}\right]$, $\lambda = -\frac{1}{2}$

3. $[x_1, x_2] = [1, 0]$, $\lambda = -1$

4. $[x_1, x_2] = [-1, 0]$, $\lambda = -1$



Rysunek 2.1: Wizualizacja ograniczeń równości Warunków Pierwszego Rzędu

2.1.2 Warunki Drugiego Rzędu w przypadku ograniczeń równościowych

Twierdzenie 4: Warunki Drugiego Rzędu Twierdzenia Lagrange'a.
Niech $f : \mathbb{D} \subset \mathbb{R}^n \rightarrow \mathbb{R}$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $h \in C^2$. *Niech istnieją takie* x *oraz* λ, *że:*

1. $\nabla_f(x) + \lambda^T \mathbf{D}h(x) = 0$, *oraz*
2. $\forall_{z \in T(x), z \neq 0}$ *mamy* $z^T H_{L(x)} z > 0$

Wtedy punkt x *nazywamy minimum funkcji* f, *związane* $h(x) = 0$. $T(x)$ *nazywamy przestrzenią styczną, tj:* $T(x) = \{z \in \mathbb{R}^n : z^T \mathbf{D}h(x) = 0\}$. *W przypadku, gdy* $m = 1$, *mamy* $T(x) = \{z \in \mathbb{R}^n : z^T \nabla_h(x) = 0\}$

Przykład 2. Rozważmy 4 rozwiązania, które uzyskaliśmy z FOC. 1. $[x_1, x_2] = \left[0, \frac{1}{\sqrt{2}}\right]$, $\lambda = -\frac{1}{2}$

$$z : z^T \nabla_h(x) = [z_1, z_2][2x_1, 4x_2]^T = [z_1, z_2]\left[0, \frac{4}{\sqrt{2}}\right] = 0$$

$$z_1 0 + z_2 \frac{4}{\sqrt{2}} = 0 \Rightarrow \mathbf{z} = [\alpha, 0]$$

$$[\alpha, 0]^T H_f([x_1, x_2])[\alpha, 0] = [\alpha, 0]^T \begin{bmatrix} 2+2\lambda & 0 \\ 0 & 2+4\lambda \end{bmatrix} [\alpha, 0] = [\alpha, 0]^T \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} [\alpha, 0] = \alpha^2 > 0$$

2. $[x_1, x_2] = \left[0, -\frac{1}{\sqrt{2}}\right]$, $\lambda = -\frac{1}{2}$
3. $[x_1, x_2] = [1, 0]$, $\lambda = -1$

$$z : z^T \nabla_h(x) = [z_1, z_2][2x_1, 4x_2]^T = [z_1, z_2]\,[1, 0] = 0$$

$$z_1 1 + z_2 0 = 0 \Rightarrow \mathbf{z} = [0, \alpha]$$

$$[0, \alpha]^T H_f([x_1, x_2])[0, \alpha] = [0, \alpha]^T \begin{bmatrix} 2+2\lambda & 0 \\ 0 & 4+\lambda \end{bmatrix} [0, \alpha] = [0, \alpha]^T \begin{bmatrix} 0 & 0 \\ 0 & -2 \end{bmatrix} [0, \alpha] = -2\alpha^2 < 0$$

4. $[x_1, x_2] = [-1, 0]$, $\lambda = -1$

2.1.3 Warunki Pierwszego Rzędu w przypadku ograniczeń *nie*równościowych

Twierdzenie 5: Warunki Karush-Kuhn-Tucker, KKT.
Niech $f : \mathbb{D} \subset \mathbb{R}^n \to \mathbb{R}$, $f \in C^1$ będzie funkcją celu, oraz $h : \mathbb{R}^n \to \mathbb{R}^m$, $h \in C^1$ oraz $g : \mathbb{R}^n \to \mathbb{R}^p$, $g \in C^1$ będą ograniczeniami. Jeżeli x jest ekstremum, to istnieje taki zestaw $\lambda = (\lambda_1, \ldots, \lambda_n)$ oraz $\mu = (\mu_1, \ldots, \mu_n)$ że:

1. ***Stationarity condition**: $\nabla_f(x) + \sum_{i=1}^m \lambda_i \nabla_{h_i(x)} + \sum_{i=1}^p \mu_i \nabla_{g_i(x)} = 0$*
2. ***Primal feasibility**: $\forall_{i=1,\ldots,m} h_i(x) = 0$ and $\forall_{i=1,\ldots,p} g_i(x) \leqslant 0$*
3. ***Dual feasibility**: $\forall_{i=1,\ldots,p} \mu_i \geqslant 0 \leftarrow$ dla minimum*
4. ***Complementary slackness**: $\forall_{i=1,\ldots,p} \mu_i g_i(x) = 0$*

Przykład 3. $f(x) = x_1^2 + x_2^2$ ograniczona przez $[x_1, x_2] : g(x) = x_1^2 + 2x_2^2 - 1 \leqslant 0$

Z FOC mamy:

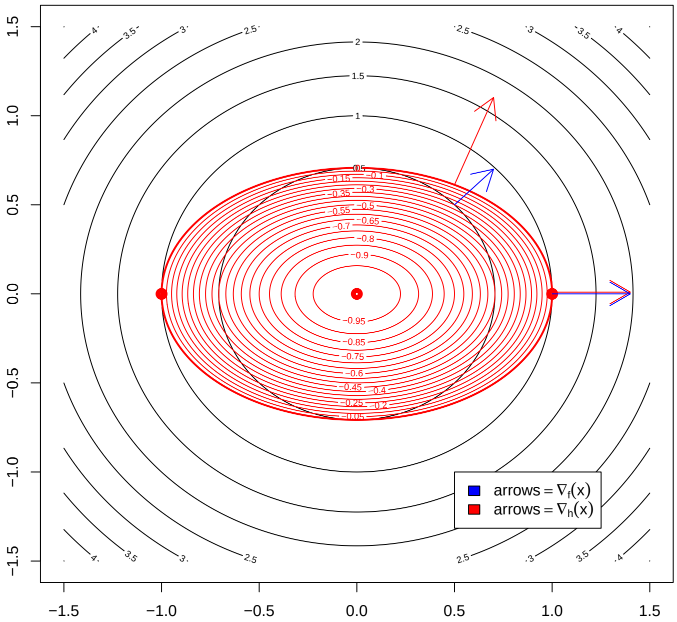
$$[2x_1, 2x_2] + \mu[2x_1, 4x_2] = 0$$

$$\begin{cases} 2x_1 + \mu 2x_1 = 0 \Rightarrow x_1(1+\lambda) = 0 \\ 2x_2 + \mu 4x_2 = 0 \Rightarrow x_2(1+2\lambda) = 0 \end{cases}$$

Co daje nam 4 rozwiązania:

1. $[x_1, x_2] = \left[0, \frac{1}{\sqrt{2}}\right]$, $\mu = -\frac{1}{2}$ **Dual feasibility**
2. $[x_1, x_2] = \left[0, -\frac{1}{\sqrt{2}}\right]$, $\mu = -\frac{1}{2}$ **Dual feasibility**

3. $[x_1, x_2] = [1, 0]$, $\mu = -1$ **Dual feasibility**
4. $[x_1, x_2] = [-1, 0]$, $\mu = -1$ **Dual feasibility**
5. $[x_1, x_2] = [0, 0]$, $\mu = 0$



Rysunek 2.2: Wizualizacja ograniczeń nierówności Warunków Pierwszego Rzędu

2.1.4 Warunki Drugiego Rzędu w przypadku ograniczeń nierównościowych

Twierdzenie 6: Twierdzenie SOC (Twierdzenie KKT).
Niech $f : \mathbb{D} \subset \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^m$, $h \in C^2$, $g : \mathbb{R}^n \to \mathbb{R}^p$, $g \in C^2$. *Niech istnieją takie* x, λ *oraz* μ, *że: 1.* $\nabla_f(x) + \lambda^T \mathbf{D}h(x) + \mu^T \mathbf{D}g(x) = 0$, *oraz 2.* $\forall_{z \in T(x), z \neq 0}$ *mamy* $z^T H_{L(x)} z > 0$

Wtedy punkt x *nazywamy minimum funkcji* f, *związane* $h(x) = 0$.

$T(x)$ *nazywamy przestrzenią styczną, tj:* $T(x) = \{z \in \mathbb{R}^n : z^T \mathbf{D}h(x) = 0\}$. *W przypadku, gdy* $m = 1$, *mamy* $T(x) = \{z \in \mathbb{R}^n : z^T \nabla_h(x) = 0\}$