Optymalizacja zaawansowana
Wykład 5

Jakub Growiec

SGH

Semestr zimowy 2019-20

- W optymalizacji często występują problemy maksymalizacji funkcji na pewnej krzywej, powierzchni, hiperpowierzchni, itp., generalnie zbiorze brzegowym ($C = \partial C$), o pustym wnętrzu.

- **Przypadek dwuwymiarowy:** krzywą wyznacza równanie $\phi(x, y) = 0$. Kiedy możemy przejść od postaci uwikłanej do postaci jawnej $y = f(x)$?
- Dla niektórych punktów może to być niemożliwe (tam, gdzie styczna jest pionowa).
- Czasem przejście jest niejednoznaczne.
- Przykład: $\phi(x, y) = x^2 + y^2 - 1 = 0$.

Twierdzenie o funkcjach uwikłanych (2 wymiary)

- Niech $U, V \subset \mathbb{R}$ będą otwarte, a funkcja $\phi : U \times V \to \mathbb{R}$ będzie klasy C^1.
- Niech $(\bar{x}, \bar{y}) \in U \times V$ spełnia $\phi(\bar{x}, \bar{y}) = 0$ oraz $\frac{\partial \phi}{\partial y}(\bar{x}, \bar{y}) \neq 0$.
  Wtedy:
- istnieją zbiory otwarte $U_{\bar{x}} \ni \bar{x}, V_{\bar{y}} \ni \bar{y}$,
- istnieje funkcja f klasy C^1, spełniająca $f : U_{\bar{x}} \to V_{\bar{y}}$, taka, że dla każdego $(x, y) \in U_{\bar{x}} \times V_{\bar{y}}$, $\phi(x, y) = 0$ jest równoważne $y = f(x)$.
- Ponadto $f'(\bar{x}) = -\frac{\frac{\partial \phi}{\partial x}(\bar{x},\bar{y})}{\frac{\partial \phi}{\partial y}(\bar{x},\bar{y})}$.

Pochodna funkcji złożonej (ang. chain rule):

- Niech $g(x) = f(u_1(x), ..., u_n(x))$. Wtedy:

$$g'(x) = \sum_{i=1}^{n} \frac{\partial f}{\partial u_i}(x) \cdot u_i'(x).$$

Dowód ostatniego punktu twierdzenia o funkcjach uwikłanych:

- Mamy $\phi(x, f(x)) = 0$ dla każdego $x \in U_{\bar{x}}$.
- Zatem również $\frac{d\phi}{dx}(x, f(x)) = 0$ na $U_{\bar{x}}$.
- Z pochodnej funkcji złożonej mamy:

$$\frac{d\phi}{dx}(x, f(x)) = \frac{\partial \phi}{\partial x}(x, y) + \frac{\partial \phi}{\partial y}(x, y) f'(x).$$

- Przekształcając i wstawiając $(\bar{x}, \bar{y})$ otrzymujemy:

$$f'(\bar{x}) = -\frac{\frac{\partial \phi}{\partial x}(\bar{x}, \bar{y})}{\frac{\partial \phi}{\partial y}(\bar{x}, \bar{y})}.$$

- **Przypadek wielowymiarowy:** podprzestrzeń / hiperpowierzchnię wyznacza równanie $\Phi(x, y) = 0$, gdzie $x = (x_1, ..., x_p)$, $y = (y_1, ..., y_n)$ oraz $\Phi : \mathbb{R}^{n+p} \to \mathbb{R}^n$.
- Mamy więc n równań, $n + p$ zmiennych:

$$\Phi(x, y) = \begin{bmatrix} \Phi_1(x_1, ..., x_p; y_1, ..., y_n) \\ \vdots \\ \Phi_n(x_1, ..., x_p; y_1, ..., y_n) \end{bmatrix} = 0.$$

- Kiedy możemy przejść od postaci uwikłanej do postaci jawnej $y = F(x)$, gdzie $F : \mathbb{R}^p \to \mathbb{R}^n$? Wyrazić wszystko w kategoriach x-ów?

Twierdzenie o funkcjach uwikłanych (wielowymiarowe)

- Niech $U \subset \mathbb{R}^p$, $V \subset \mathbb{R}^n$ będą otwarte, a funkcja $\Phi : U \times V \to \mathbb{R}^n$ będzie klasy C^1.
- Niech $(\bar{x}, \bar{y}) \in U \times V$ spełnia $\Phi(\bar{x}, \bar{y}) = 0$ oraz $\det(\frac{\partial \Phi}{\partial y}(\bar{x}, \bar{y})) \neq 0$ (macierz pochodnych cząstkowych względem y jest odwracalna). **Wtedy:**
- istnieją zbiory otwarte $U_{\bar{x}} \ni \bar{x}$, $V_{\bar{y}} \ni \bar{y}$,
- istnieje funkcja F klasy C^1, spełniająca $F : U_{\bar{x}} \to V_{\bar{y}}$, taka, że dla każdego $(x, y) \in U_{\bar{x}} \times V_{\bar{y}}$, $\Phi(x, y) = 0$ jest równoważne $y = F(x)$.
- Ponadto $F'(\bar{x}) = -\left(\frac{\partial \phi}{\partial y}(\bar{x}, \bar{y})\right)^{-1} \cdot \frac{\partial \Phi}{\partial x}(\bar{x}, \bar{y})$.

- **Przypomnienie twierdzenia.** Niech $C \subset \mathbb{R}^n$, a funkcja $f : C \to \mathbb{R}$. Rozważmy problem

$$(P) \quad \max_{x \in C} f(x).$$

Jeśli punkt $\bar{x} \in C$ jest rozwiązaniem problemu (P) oraz f jest różniczkowalna w $\bar{x}$, to

$$\bigtriangledown f(\bar{x}) \in N_C(\bar{x}).$$

- **Uwaga.** Warto umieć szybko obliczać $N_C(\bar{x})$ dla przypadku, gdy zbiór C jest zadany ograniczeniem w formie równości, np. $\varphi(x) = 0$.
- Wtedy łatwo będzie nam znajdować warunki konieczne optymalizacji z tego rodzaju ograniczeniami.
- Można to osiągnąć poprzez **mnożniki Lagrange'a**, o ile tylko spełnione będą **warunki regularności**.

Warunki regularności:

- Niech
  $C = \{(x_1, ..., x_n) \in \mathbb{R}^n : g^1(x_1, ..., x_n) = ... = g^k(x_1, ..., x_n) = 0\}$.
- Mamy n zmiennych oraz układ $k \leq n$ (zwykle $k < n$) warunków ograniczających w formie równości.
- Punkt $\bar{x} \in C$ jest **regularny**, jeśli wektory $\triangledown g^1(\bar{x}), ..., \triangledown g^k(\bar{x})$ są liniowo niezależne.
- Inaczej, jeśli rząd macierzy

$$\begin{bmatrix} \triangledown g^1(\bar{x}) \\ \vdots \\ \triangledown g^k(\bar{x}) \end{bmatrix}$$

  wynosi k (jest maksymalny).

Sprawdzić regularność punktów spełniających układ równań $g(x) = 0$:

- **Przykład 1.** Niech $k = 2, n = 3$. Niech $g^1(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2$ oraz $g^2(x_1, x_2, x_3) = x_1 + x_2 + x_3$.
- **Przykład 2.** Niech $k = 2, n = 3$. Niech $g^1(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 - 1$ oraz $g^2(x_1, x_2, x_3) = x_1 + x_2 + x_3$.
- **Przykład 3.** Niech $k = 2, n = 3$. Niech $g^1(x_1, x_2, x_3) = x_1^2 + x_2 - 1$ oraz $g^2(x_1, x_2, x_3) = x_1 + x_2 + x_3 + 2$.

- Jeśli zbiór $C = \{(x_1, ..., x_n) \in \mathbb{R}^n : g^1(x_1, ..., x_n) = ... = g^k(x_1, ..., x_n) = 0\}$ jest niepusty, przy spełnieniu warunków regularności dla wszystkich $x \in C$ daje się on „sparametryzować" za pomocą $n - k$ zmiennych.
- Jest to wniosek z twierdzenia o funkcjach uwikłanych.
- Taki zbiór C nazywamy $n - k$ wymiarową **rozmaitością** w $\mathbb{R}^n$ (ang. *manifold*).
- W szczególności jeśli mamy tylko 1 warunek ograniczający, $k = 1$, wtedy taki zbiór C jest $n - 1$ wymiarową rozmaitością (= **hiperpowierzchnią**).

Stożek normalny i styczny do rozmaitości:

- Niech zbiór $C = \{(x_1, ..., x_n) \in \mathbb{R}^n : g^1(x_1, ..., x_n) = ... = g^k(x_1, ..., x_n) = 0\}$ będzie niepusty i regularny. Wtedy dla każdego $\bar{x} \in C$

$$N_C(\bar{x}) = \left\{ \sum_{i=1}^{k} \lambda_i \bigtriangledown g^i(\bar{x}), \ \lambda_i \in \mathbb{R} \right\} = L\left(\bigtriangledown g^1(\bar{x}), ..., \bigtriangledown g^k(\bar{x})\right)$$

oraz

$$T_C(\bar{x}) = \{h \in \mathbb{R}^n :< \bigtriangledown g^i(\bar{x}), h >= 0, \ i = 1, ..., k\}.$$

Warunki konieczne optymalizacji z ograniczeniami w formie równości

- **Twierdzenie.** Niech $g^1, ..., g^k : \mathbb{R}^n \to \mathbb{R}$, gdzie $k \leq n$, będą klasy C^1. Niech zbiór $C = \{(x_1, ..., x_n) \in \mathbb{R}^n : g^1(x_1, ..., x_n) = ... = g^k(x_1, ..., x_n) = 0\}$ będzie niepusty i regularny.
- Rozważmy problem
$$(P) \quad \max_{x \in C} f(x).$$
- Jeśli $\bar{x} \in C$ jest rozwiązaniem (P) oraz f jest różniczkowalna w $\bar{x}$, wtedy istnieją **mnożniki Lagrange'a** $\lambda_1, ..., \lambda_k \in \mathbb{R}$ takie, że:
$$\triangledown f(\bar{x}) = \sum_{i=1}^{k} \lambda_i \triangledown g^i(\bar{x}).$$

Warunki konieczne optymalizacji z ograniczeniami w formie równości

- **Twierdzenie.** Niech $g^1, ..., g^k : \mathbb{R}^n \to \mathbb{R}$, gdzie $k \leq n$, będą klasy C^1. Niech zbiór $C = \{(x_1, ..., x_n) \in \mathbb{R}^n : g^1(x_1, ..., x_n) = ... = g^k(x_1, ..., x_n) = 0\}$ będzie niepusty i regularny.
- Rozważmy problem
$$(Q) \quad \min_{x \in C} f(x).$$
- Jeśli $\bar{x} \in C$ jest rozwiązaniem (Q) oraz f jest różniczkowalna w $\bar{x}$, wtedy istnieją **mnożniki Lagrange'a** $\lambda_1, ..., \lambda_k \in \mathbb{R}$ takie, że:
$$\bigtriangledown f(\bar{x}) = \sum_{i=1}^{k} \lambda_i \bigtriangledown g^i(\bar{x}).$$

- **Definicja.** Funkcja Lagrange'a (lagrangian) to funkcja $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}$ postaci

$$\mathcal{L}(x; \lambda) = f(x) - \sum_{i=1}^{k} \lambda_i g^i(x).$$

- Łatwo zauważyć, że

$$\bigtriangledown f(\bar{x}) = \sum_{i=1}^{k} \lambda_i \bigtriangledown g^i(\bar{x}) \iff \bigtriangledown \mathcal{L}(\bar{x}; \lambda) = 0$$

(gradient lagrangianu liczymy różniczkując po x, nie po λ).

Warunki drugiego rzędu dla optymalizacji z ograniczeniami w formie równości:

- **Intuicja.** Działamy analogicznie, jak w przypadku optymalizacji bez ograniczeń, z tym że interesuje nas kierunek zmian f tylko „wzdłuż ograniczenia", czyli **w ramach stożka stycznego** $T_C(\bar{x})$.
  (a) Przypadek łatwiejszy (czasem wystarcza, ale nie zawsze)
- Jeśli f jest wypukła w pewnym zbiorze otwartym $U_{\bar{x}} \ni \bar{x}$, to $\bar{x}$ jest minimum lokalnym.
- Jeśli f jest wklęsła w pewnym zbiorze otwartym $U_{\bar{x}} \ni \bar{x}$, to $\bar{x}$ jest maksimum lokalnym.
  (b) Przypadek trudniejszy (kiedy przypadek łatwiejszy nie wystarcza)
- Kiedy ww. warunki nie zachodzą, trzeba się ograniczyć do zbioru $U_{\bar{x}} \cap T_C(\bar{x})$.
- Wtedy badamy określoność formy kwadratowej

$$h^T D^2 f(\bar{x}) h, \qquad \forall h \in T_C(\bar{x}).$$

- Jeśli wyrażenie jest stale nieujemne, to mamy minimum lokalne.
- Jeśli wyrażenie jest stale niedodatnie, to mamy maksimum lokalne.