Metody Rozmyte i Algorytmy Ewolucyjne

mgr inż. Piotr Kaczyński

Wydział Matematyczno-Przyrodniczy
Szkoła Nauk Ścisłych
Uniwersytet Kardynała Stefana Wyszyńskiego

Podstawy optymalizacji

Plan prezentacji

Podstawy matematyczne

Własności zadań optymalizacji
- Rodzaje zadań
- Metody optymalizacji
- Środowisko dynamiczne

Ograniczenia
- Eliminacja ograniczeń
- Metody funkcji kary

Wstęp

- Dana jest metryczna przestrzeń poszukiwań
$$\Omega = (U, |\cdot|)$$
gdzie U jest zbiorem wartości, a $|\cdot|$ pewną metryką,
- Dany jest również pewien podzbiór $D \subseteq U$,
- W zadaniu dana jest również pewna funkcja $F(x)\colon U \to \mathbb{R}$ zwana funkcją celu,
- Zadanie optymalizacji polega na znalezieniu takiego $\hat{x} \in D$, że
$$\hat{x} = \arg\min_{x \in D} F(x)$$
- Zadanie minimalizacji można sprowadzić do zadania maksymalizacji funkcji $-F(x)$.

Pochodne

Definition
Jeżeli funkcja $F\colon \mathbb{R}^n \to \mathbb{R}$ posiada wszystkie pochodne cząstkowe, to wektor wierszowy
$$\nabla F = \begin{bmatrix} \frac{\partial F}{\partial x_1} & \dots & \frac{\partial F}{\partial x_n} \end{bmatrix}$$
nazywamy gradientem funkcji F.

Definition
Jeśli funkcja $F\colon \mathbb{R}^n \to \mathbb{R}$ posiada wszystkie pochodne cząstkowe do drugiego rzędu włącznie, to macierz
$$\nabla^2 F = \begin{bmatrix} \frac{\partial^2 F}{\partial x_i \partial x_j} \end{bmatrix}$$
nazywamy macierzą Hessego(Hesjanem) funkcji F.

Rozwinięcie w szereg Taylora

Definition

Rozwinięciem w szereg Taylora funkcji $F\colon \mathbb{R} \to \mathbb{R}$ wokół punktu x_0 nazywamy szereg

$$f(x_0) + \sum_{i=1}^{\infty} \frac{1}{i!} \frac{d^i F}{dx^i}(x_0)(x - x_0)^i$$

Przykład

Funkcję $F\colon \mathbb{R}^n \to \mathbb{R}$ można przybliżyć pierwszymi dwoma wyrazami szeregu

$$F(x) \approx F(x_0) + \nabla F(x_0)(x - x_0) + \frac{1}{2}\nabla^2 F(x_0)(x - x_0)^2$$

Warunki konieczne optymalności

Theorem

$$\nabla F(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) = 0$$
$$\forall i = 1, \ldots, m \quad \lambda_i g_i(x) = 0$$
$$\forall i = 1, \ldots, m \quad \lambda_i \geqslant 0$$
$$\forall i = 1, \ldots, m \quad g_i(x) \leqslant 0$$

Podział ogólny

- Optymalizacja ciągła ($U = \mathbb{R}^n$)
  - Zadania wypukłe,
  - Zadania optymalizacji globalnej,
- Optymalizacja dyskretna
  - Zadania kombinatoryczne (zmienne binarne),
  - Zadania mieszane,

Właściwości funkcji celu

- Zadania w $\mathbb{Z}^n$ będące dyskretną wersją z $\mathbb{R}^n$
  - Nałożenie ograniczeń całkowitoliczbowych na zadania sformułowane w $\mathbb{R}^n$,
  - Można próbować rozwiązać metodami relaksacji ograniczenia $x \in \mathbb{Z}^n$,
  - Rozwiązanie problemu ciągłego stanowi oszacowanie zadania dyskretnego,
  - Prosty problem ciągły nie gwarantuje prostego problemu dyskretnego!
- Liniowość (afiniczność)
  - Funkcja F jest liniowa, jeśli

    $$F(x) = a^T x + b$$

    gdzie a jest n-wymiarowym wektorem, b - stałą.
  - Rozwiązanie zadania leży zawsze na brzegu zbioru rozwiązań dopuszczalnych D.

Właściwości funkcji celu II

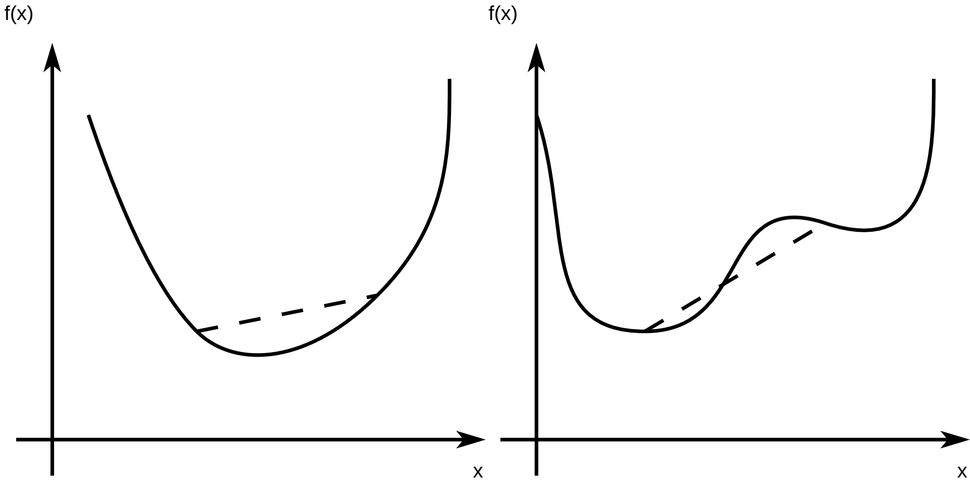
- Wypukłość
  - Funkcja wypukła spełnia warunek

    $$\forall x_1, x_2 \in D \quad \lambda F(x_1) + (1-\lambda)F(x_2) > F(\lambda x_1 + (1-\lambda)x_2)$$

    gdzie $0 < \lambda < 1$ oraz zbiór D jest wypukły,
  - Istnieje wtedy dokładnie jedno minimum funkcji celu,
  - Bardzo upraszcza poszukiwanie optimum.
- Różniczkowalność
  - Funkcja posiada pochodne cząstkowe pierwszego i/lub drugiego rzędu,
  - Jedna z najbardziej przydatnych własności,

Wypukłość funkcji celu

Przykład



Właściwości funkcji celu III

- Warunek Lipshitza
  - Warunek na szybkość zmienności funkcji,
  - Funkcja spełnia ten warunek, jeśli istnieje takie $L < \infty$, że

    $$\forall x_1, x_2 \quad |F(x_1) - F(x_2)| \leqslant L|x_1 - x_2|$$

  - Rzadko kiedy jest znany w praktyce,
- Właściwość dekompozycji
  - Funkcja mająca tą własność jest złożeniem wielu funkcji, z których każda zależy tylko od części zmiennych

    $$F(x) = \mathcal{F}(f_i(x_{(i)}))$$

    gdzie $f_i(x_{(i)})$ są funkcjami, których argument zawiera część wektora x,
  - Wystarczy zminimalizować f_i i na końcu funkcję $\mathcal{F}$,
  - To upraszcza zadanie, każda z funkcji ma mniej zmiennych niż zawiera wektor x.

Wiele minimów lokalnych

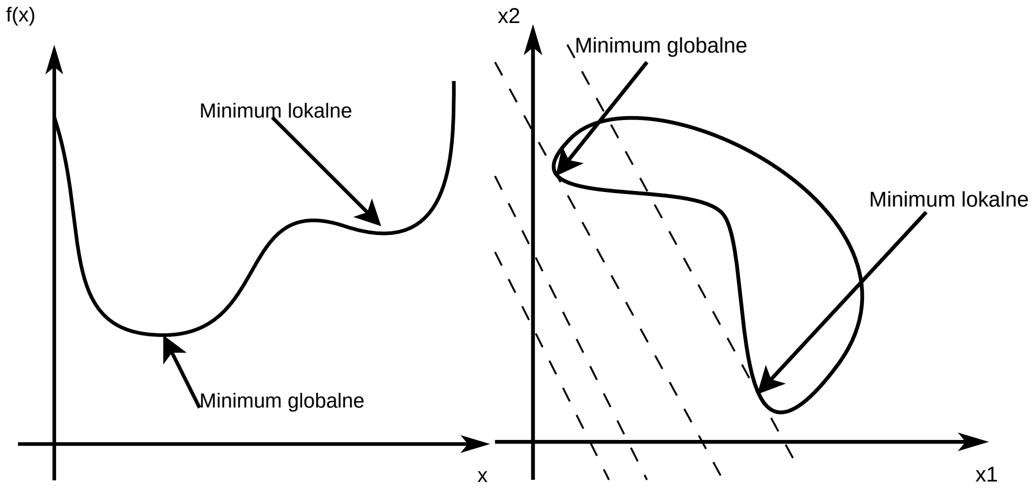
- Minimum lokalne to punkt x, dla którego

  $$\exists \delta > 0 \quad \forall r < \delta \quad \forall y \in K(x, r) \cap D \qquad F(x) \leqslant F(y)$$

- Może się zdarzyć, że istnieje spójny zbiór o jednakowej wartości,
- Zbiór taki traktujemy jako jeden punkt i nazywamy minimum niewłaściwym
- Istnienie minimów lokalnych wynika z
  - Niewypukłości funkcji celu,
  - Niewypukłości zbioru rozwiązań dopuszczalnych,

Wiele minimów lokalnych II

Przykład



Zbiory przyciągania minimów lokalnych

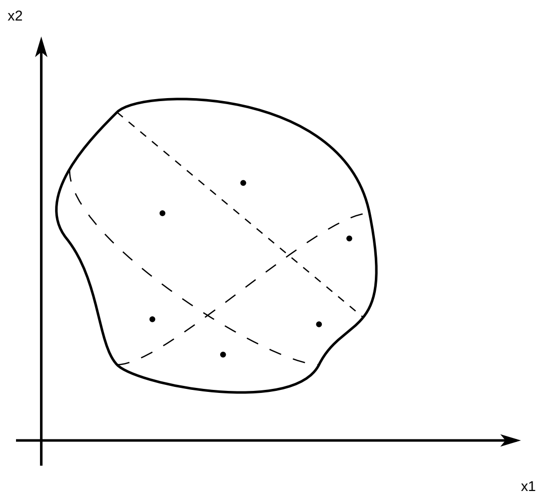
- Pojęcie związane z algorytmami optymalizacji lokalnej,
- Rozważmy odwzorowanie

$$y(x) = \begin{cases} z^* \text{ takie, że} \\ \forall z \in K(x,r) \cap D\ F(z^*) \leqslant F(z) \text{ oraz } F(z^*) < F(x) \\ x \text{ w przeciwnym przypadku} \end{cases}$$

- Rozważmy ciągi punktów $x_{i+1} = y^{(i)}(x_i)$ (i-krotne złożenie funkcji y),
- Zbiór wszystkich punktów, które są elementami początkowymi tego ciągu zbieżnego do minimum lokalnego $\hat{x}$ nazywamy obszarem przyciągania tego minimum,
- Relacja należenia do obszaru przyciągania danego minimum jest relacją równoważności

Zbiory przyciągania minimów lokalnych

Przykład



Ograniczenia funkcji celu

- Zbiór dopuszczalny definiujemy za pomocą ograniczeń postaci

$$g_i(x) \leqslant 0$$
$$h_j(x) = 0$$

- Ograniczenia g_i nazywamy nierównościowymi,
- Ograniczenia h_j nazywamy równościowymi
- Najważniejsze (najprostsze) typy ograniczeń
  - Kostkowe
  $$l_i \leqslant x_i \leqslant u_i, \quad l_i, u_i \in \mathbb{R}$$
  - Liniowe
  $$g_i(x) = a^T x + b, \quad a \in \mathbb{R}^n, b \in \mathbb{R}$$
  - Wymienione typy ograniczeń zawsze tworzą wypukły obszar rozwiązań dopuszczalnych

Ilość wymiarów

- Im większa ilość wymiarów, tym trudniejsze zadanie,
- Zarówno dla zagadnień dyskretnych jak i ciągłych,

Przykład

Rozważmy metodę optymalizacji polegającą na „trafianiu”. Chcemy trafić w koło wpisane w kwadrat. Prawdopodobieństwo tego wynosi $\frac{\pi}{4}$.
Jeśli przyjmiemy wymiar zadania o jeden większy (sześcian i kula) to prawdopodobieństwo to wyniesie $\frac{\pi}{6}$.
Uogólniając na coraz większą liczbę wymiarów, prawdopodobieństwo to będzie coraz mniejsze.

Przykład

Jak zmieni się przestrzeń przeszukiwań dla zadania binarnego?

Zadania ciągłe

- Dla zadań liniowych - metoda sympleksów,
- Warunki konieczne optymalności,
- Metody bezgradientowe
  - Metoda sympleksu Neldera-Meada,
- Metody gradientowe
  - Metoda największego spadku,
  - Metoda gradientów sprzężonych,
- Metody (quasi-)Newtonowskie
  - Metoda Newtona,
  - Metoda DFP,
  - Metoda BFGS,

Schemat metod I i II rzędu

1. Wyznacz rozwiązanie początkowe x_i, $i = 0$
2. Mając $\nabla F(x_i)$ (oraz $\nabla^2 F(x_i)$) wyznacz odpowiedni kierunek d,
3. Zminimalizuj (dokładnie lub niedokładnie) funkcję jednej zmiennej
$$\bar{x} = \min_{\alpha \in \mathbb{R}} F(x_i + \alpha d)$$
4. Podstaw $x_{i+1} = \bar{x}$, oraz $i = i + 1$,
5. Jeśli nie jest spełniony warunek STOP, to wróć do kroku 2

Zadania dyskretne

- Brak szybkich i pewnych metod optymalizacji,
- Metody dokładne
  - Przeszukiwanie całego zbioru rozwiązań dopuszczalnych,
  - Metoda podziału i oszacowań,
  - Metody niedeterministyczne

Metody optymalizacji globalnej

- Metody poszukiwania optimum globalnego,
- Metody wielostartowe,
- Optymalizacja lokalna z momentem,
- Poszukiwanie z tabu,
- Metody niedeterministyczne
  - Metoda Monte-Carlo (wielokrotne losowanie przypadkowych punktów),
  - Metoda symulowanego wyżarzania (pewne prawdopodobieństwo pójścia w „złym” kierunku),
  - Algorytmy genetyczne

Sformułowanie problemu

- Optymalizacja funkcji niestacjonarnej przy niestacjonarnych ograniczeniach

$$\hat{x}(t) = \arg \min_{x \in D(t)} F(x, t)$$

  gdzie t oznacza czas,
- Rozwiązaniem nie jest punkt, ale funkcja zależna od t,
- Warto zwrócić uwagę, że zbiór rozwiązań dopuszczalnych też może się zmieniać w czasie,
- Zadanie trudniejsze niż optymalizacja statyczna

Metody rozwiązywania zadań niestacjonarnych

- Konieczne zastosowanie adaptacji,
- Podstawowa metoda: restart algorytmu po zajściu zmiany,
- Jak zmiany nie są znane - restart jak najczęściej,
- Jeśli położenia minimów zmieniają się nieznacznie — start od punktu poprzednio znalezionego

Eliminacja ograniczeń

- Dla ograniczeń równościowych, które da się przekształcić do postaci

$$x_{(i)} = \psi(x_{(j)})$$

  gdzie $x_{(i)}$ oraz $x_{(j)}$ są pewnymi rozłącznymi podwektorami wektora x, stosujemy podstawienie do funkcji celu,
- Sytuacja bardzo rzadka w praktyce,

Przykład

Dla zbioru ograniczeń liniowych postaci

$$A^T x + b = [A_B\ A_N][x_B\ x_N]^T + b = 0$$

gdzie A_B to niesingularna macierz bazowa, otrzymamy podstawienie

$$x_B = -A_B^{-1}(A_N x_N + b)$$

Zewnętrzna funkcja kary

- Polega na modyfikacji funkcji celu,
- Funkcja celu w obszarze dopuszczalnym nie jest zmieniana,
- Nakładana jest „kara” za przekroczenie ograniczeń,
- Funkcja celu musi być określona poza obszarem dopuszczalnym,

Definition

Zewnętrzną funkcją kary nazywamy każdą funkcję $f\colon U \to \mathbb{R}$ taką, że

$$\begin{aligned} f_i^p(x) &= 0 \quad \text{jeśli} g_i(x) \leqslant 0 \\ f_i^p(x) &> 0 \quad \text{jeśli} g_i(x) > 0 \end{aligned}$$

Przykład

Rysunek

Zewnętrzna funkcja kary II

- Co zrobić, aby znaleźć szukane minimum?
- Stosuje się iteracyjną zmianę funkcji kary,

$$f_i^{p,(k+1)}(x) \geqslant f_i^{p,(k)}(x)$$
$$\lim_{k\to\infty} f_i^{p,(k)}(x) = \infty \text{ dla } g_i(x) > 0$$

- Szczególny przypadek: „kara śmierci”

$$f_i^p(x) = \infty \quad x \notin D$$

- Zmiany funkcji kary kończymy, jak znalezione optimum znajduje się w obszarze ograniczeń,

Wewnętrzna funkcja kary

- Również polega na modyfikacji funkcji celu,
- Funkcja celu jest zmieniana w obszarze dopuszczalnym,
- Nakładana jest „kara” za zbliżanie się do ograniczeń,
- Stosowana, gdy funkcja celu nie jest zdefiniowana poza obszarem dopuszczalnym,

Definition

Wewnętrzną funkcją kary nazywamy każdą funkcję $f\colon D \to \mathbb{R}$ taką, że

$$f_i^p(x) \geqslant 0$$
$$\lim_{g_i(x)\to 0^-} f_i^p(x) = \infty$$

Przykład

Rysunek

Wewnętrzna funkcja kary II

- Co zrobić, aby znaleźć szukane minimum?
- Stosuje się iteracyjną zmianę funkcji kary,

$$f_i^{p,(k+1)}(x) \leqslant f_i^{p,(k)}(x)$$
$$\lim_{k\to\infty} f_i^{p,(k)}(x) = 0 \text{ dla } g_i(x) < 0$$

- Zmiany funkcji kary kończymy, jeśli znajdowane rozwiązania nie zmieniają się,