Układy równań nieliniowych

Ogólna teoria

- W przypadku układów równań nieliniowych możemy zastosować podobne podejście jak dla pojedynczych równań.

$$f(x) = 0, \qquad f: \mathbb{R}^n \to \mathbb{R}^n$$

- Koncepcja opiera się o stworzenie modelu zmiany residuum w związku ze zmianą

$$x_{k+1} = x_k + p$$

- W związku z tym modelujemy $r(x_k + p)$

$$r(x_k + p) \approx r(x_k) + J_k p$$

Metoda Newtona

- Idea polega na znalezieniu takiego p, żeby $r(x_k + p) = 0$, tj.

$$J_k p = -r(x_k)$$

- J_k powinno być macierzą pochodnych cząstkowych (macierzą Jakobiego) w punkcie x_k

$$J_k = J(x_k) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}$$

- Lub jej aproksymacją (inne metody)

Zbieżność

- Jeżeli jesteśmy blisko rozwiązania a funkcja f jest różniczkowalna, a pochodna spełnia warunek Lipshitza, to mamy zbieżność kwadratową

$$x_{k+1} - x^* = O(\|x_{k+1} - x^*\|^2)$$

Niedokładne metody Newtona

- Istnieje klasa metod, która wyznacza kolejny krok rozwiązania p jako

$$\|r_k + J_k p_k\| \leq \eta_k \|r_k\|$$

gdzie $\eta_k < \eta$, $\eta \in [0,1]$ to tzw. ciąg wymuszający

Metoda Broydena

- Metoda Broydena jest uogólnieniem metody siecznych
- Niech $s_k = x_{k+1} - x_k,\ y_k = r(x_{k+1}) - r(x_k)$
- Konstruujemy aproksymację macierzy Jakobiego B_k, która ma spełniać tzw. warunek siecznej

$$y_k = B_{k+1} s_k$$

Wzór Broydena

- Najpopularniejszą aproksymacją jest tzw. wzór Broydena

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k) s_k^T}{s_k^T s_k}$$

Jest to najmniejsza możliwa zmiana w aproksymacji przy spełnieniu warunku siecznej.

Metody praktyczne – funkcja celu

- Metody Newtonowskie i quasi-Newtonowskie są zbieżne jedynie lokalnie.
- Aby uniknąć tego problemu, zaproponowano aby korygować długość kroku między iteracjami tj.

$$x_{k+1} = x_k + \varepsilon p$$

- Długość kroku, dobieramy tak aby poprawić wartość funkcji celu

Funkcja celu – suma kwadratów residuów

- Najpopularniejszą funkcją celu jest suma kwadratów residuów

$$m(x) = \frac{1}{2}\sum_{i=1}^{n} r_i^2(x)$$

- Z oczywistych względów pierwiastek równania jest minimum funkcji celu

Poszukiwanie kierunkowe

- Podsumowując, kolejne kroki rozwiązania znajdujemy zmniejszają wartość funkcji celu.
- Nie jest konieczne poszukiwanie dokładne minimum
- Wystarczy spełnienie tzw. warunków Wolfe'a, w uproszczeniu:
  - Nie robimy długich kroków, chyba że jest to bardzo uzasadnione
  - Jeżeli funkcja po naszym kroku dalej silnie maleje, to trzeba zrobić dłuższy krok

Problem

- Podobnie jak w jednowymiarowej metodzie Newtona problemem jest jak pochodna się zeruje.
- W tym przypadku, pochodna (gradient) funkcji celu wynosi:

$$\nabla m(x) = J(x_k)^T r(x_k)$$

- Oznacza to, że jeżeli J jest osobliwa, to dla niezerowych residuuów możliwe jest istnienie minimum funkcji celu.

Metody regionu zaufania (trust region)

- Aby uniknąć problemów ze złymi kierunkami i osobliwością zaproponowano metodę mocniej ugruntowaną w optymalizacji.
- Wyznaczamy następny krok nie na kierunku tylko wewnątrz pewnego zbioru (regionu zaufania), w którym jesteśmy pewni naszego modelu.
- Zbiór aktualizujemy z iteracji na iterację
- Jak zbliżamy się do rozwiązania przechodzi to w metodę Newtona, ale uniemożliwia ucieczkę.

Metody w Pythonie - scipy.optimize.root

- Parametr ‘method’
- *hybr* - modyfikacja metody Powella (trust region)
- *lm* - nieliniowy problem najm. kw. Metodą Levenberga-Marquardta (zasadniczo trust region)
- *df-sane* - metoda spektralna bezgradientowa (zasadniczo interpolacja
- *broyden1*, *broyden2*, *anderson*, *linearmixing*, *diagbroyden*, *excitingmixing*, *krylov* różne warianty metody Newtona

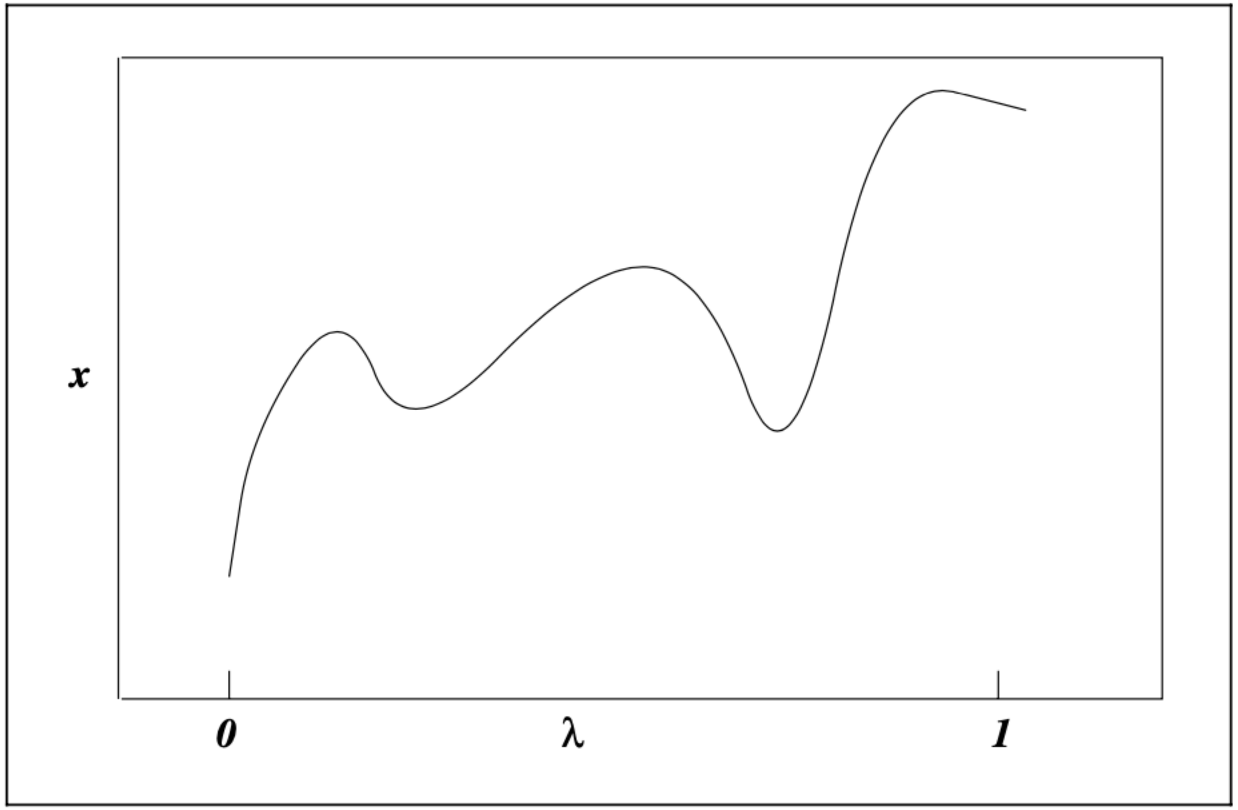
Alternatywa – homotopia/przedużanie

- Pomysł polega na stopniowym przechodzeniu od problemu prostszego do trudniejszego

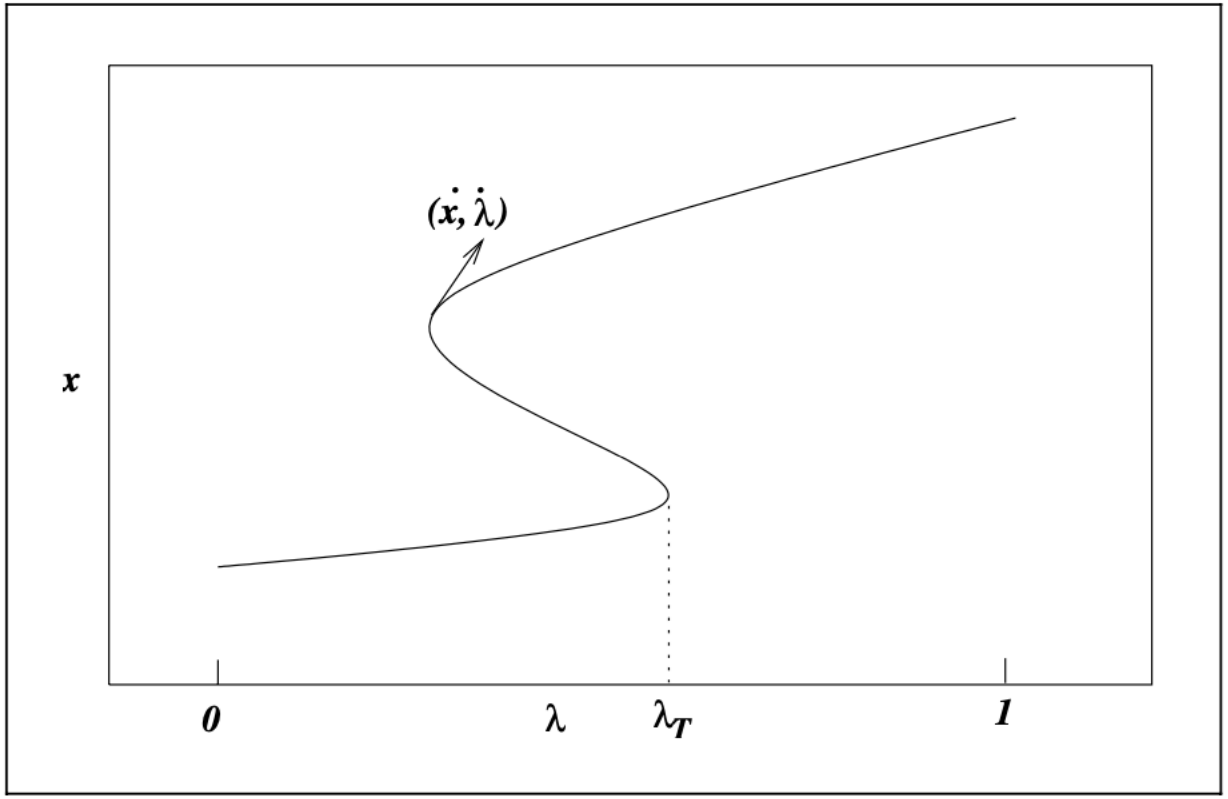
$$H(x, \lambda) = \lambda r(x) + (1 - \lambda)(x - a)$$

- $\lambda \in [0,1]$, zaczynamy od prostego problemu dla $\lambda = 0$ i stopniowo go zmieniając zmierzamy do rozwiązania

Jeśli jest dobrze



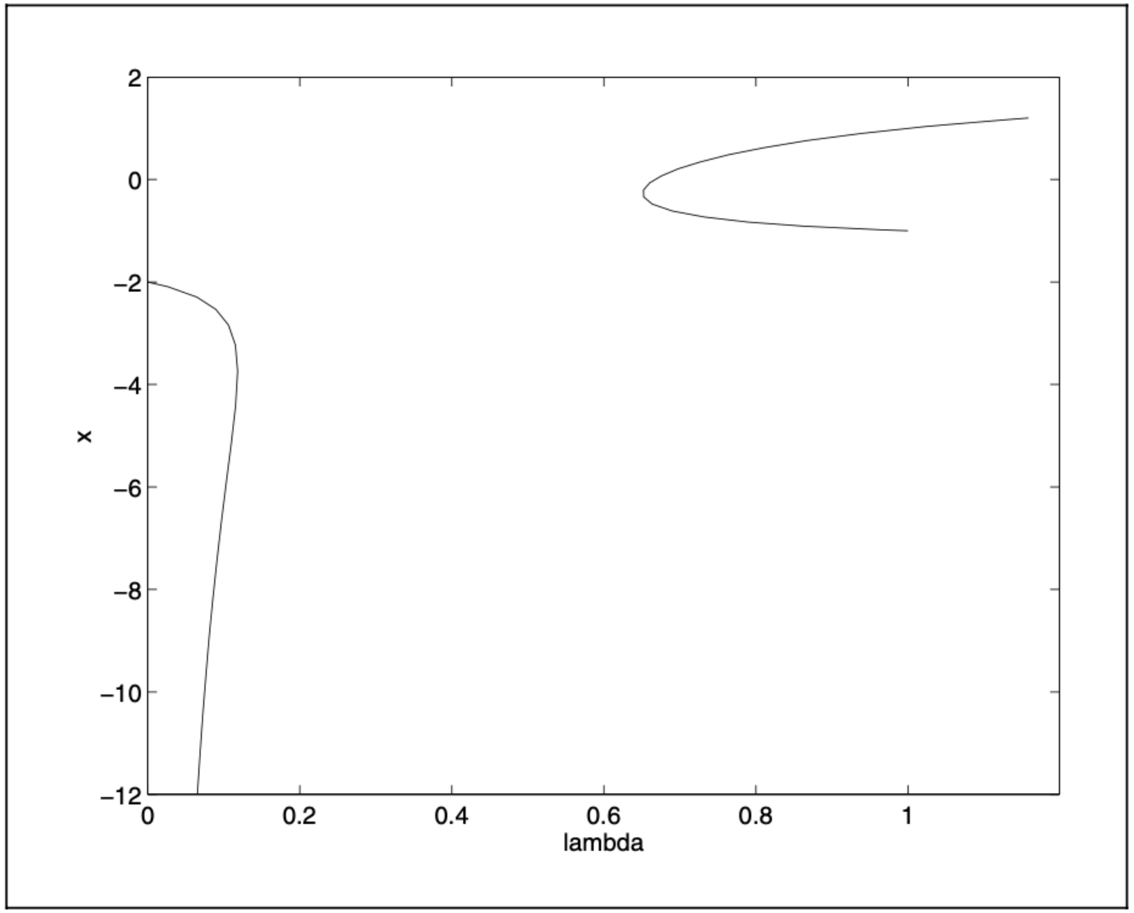
Czasem gorzej



W praktyce

- Wyznaczamy wektor styczny do homotopii
- W oparciu o niego konstruujemy albo równania różniczkowe albo algebraiczne, które trzymają nas dalej na ścieżce

Najgorzej



$$H(x, \lambda) = \lambda(x^2 - 1)$$
$$+(1 - \lambda)(x + 2)$$