

Polskie Towarzystwo Badań  
Operacyjnych i Systemowych  
Instytut Badań Systemowych  
Polskiej Akademii Nauk  
Wojskowa Akademia Techniczna

Redaktorzy:  
Zbigniew Nahorski  
Marian Chudy  
Andrzej Straszak



Warszawa 1991

POLSKIE TOWARZYSTWO  
BADAŃ OPERACYJNYCH I SYSTEMOWYCH  
INSTYTUT BADAŃ SYSTEMOWYCH  
POLSKIEJ AKADEMII NAUK  
WOJSKOWA AKADEMIA TECHNICZNA

# O P T Y M A L I Z A C J A

ZADANIA, METODY, ALGORYTMY

Redaktorzy

*Zbigniew Nahorski, Marian Chudy, Andrzej Straszak*

WARSZAWA 1991

O REGULACH ZATRZYMYWANIA ITERACYJNYCH  
ALGORYTMÓW OPTIMALIZACJI

Lesław Paradowski  
Wojskowa Akademia Techniczna  
ul. Kaliskiego 2, 00-908 Warszawa

**Streszczenie:** Sformułowano testy zatrzymywania (przerywania) procesu iteracyjnego dla nieliniowych zadań optymalizacji. Do konstrukcji reguł decyzyjnych zastosowano trzy szczególne normy wektorów, wynikające z ogólnej  $p$ -normy Hőlder'a i posiadające interpretację geometryczną: normę tzw. pierwszą  $\|x\|_1$ , euklidesową  $\|x\|_2$  oraz Czebyszewa  $\|x\|_\infty$ . Z postaci przyjętych funkcji decyzyjnych wynika kształt odpowiednich obszarów dopuszczalnych (obszary zatrzymania procesu iteracyjnego) w postaci odpowiednio: hiperwielościanu, hiperelipsoidy oraz hiperprostokątścianu. Określono miary ostrości przedstawionych reguł decyzyjnych.

### 1. Wstęp

W nieliniowych zadaniach optymalizacji proces poszukiwania ekstremum funkcji celu, lub ogólnie - funkcjonału jakości, jest z natury zorganizowany w sposób iteracyjny. Ogólnie jednak wybór testu zatrzymywania (zwanego również regułą stopowania [1] lub kryterium kończenia [2]) procesu iteracyjnego (ZPI) poszukiwania wspomnianego ekstremum jest zadaniem złożonym i stanowi swojego rodzaju sztukę [3]. W pracy [3], jako ogólną wskazówkę pomocną w tej kwestii, wymienia się wykorzystanie praktycznych ilorazów zbieżności ([3]s.432). Mianowicie, jeżeli tylko wiadomo, że algorytm jest zbieżny nadliniowo, to mała wartość praktycznego ilorazu zbieżności świadczyć będzie o zbliżaniu się do rozwiązania i może służyć jako test zatrzymywania algorytmu. Jeżeli jednak typ (charakter) zbieżności algorytmu nie jest znany, to znalezienie dobrej reguły zatrzymywania jest trudne. Ogólnie można wymienić trzy podstawowe grupy testów zatrzymywania [3].

### Reguły zatrzymywania algorytmów

1) Proste testy uniwersalne (ogólne), odznaczające się najszerszym zakresem stosowalności, jednak zawodne w bardziej złożonych przypadkach; nie wykorzystują one bliższych informacji o minimalizowanym funkcjonału (funkcji).

2) Testy specjalne, zorientowane problemowo tzn. wykorzystujące informację związaną z charakterem minimalizowanego funkcjonału oraz właściwościami algorytmu optymalizacji; im szerzej i bardziej szczegółowo uwzględniają one właściwości funkcjonału i algorytmu, tym większą odznaczają się niezawodnością.

3) Rozbudowane (złożone) testy uniwersalne, wymagające złożonych obliczeń uzupełniających; nierazdo są to rozbudowane kombinacje dwóch poprzednich testów albo testy dwuczłowe, w których test główny jest realizowany po spełnieniu testu wstępnego (test-preselktor), którym może być prosty test uniwersalny lub test specjalny.

Poniżej zostanie przedstawiona pewna ogólna reguła ZPI oraz wynikające z niej trzy warianty szczególne. Z punktu widzenia przytoczonej klasyfikacji testów zatrzymywania można ją zaliczyć do grupy prostych testów uniwersalnych z użyciem wskaźników definiowanych w przestrzeni rozwiązań, tj. przestrzeni poszukiwanych (optymalizowanych) parametrów. Regułę taką można stosować np. w zagadnieniu estymacji nieliniowej wektora parametrów  $x \in \mathbb{R}^n$  z kwadratową funkcją celu (ryzyka) w postaci uogólnionej sumy kwadratów błędów [4-6].

### 2. Sformułowanie problemu

Do utworzenia ogólnej, a zarazem względnie prostej reguły ZPI można zaproponować użycie p-normy Höldera, definiowanej dla wektora n-wymiarowego

$$x = [x_1, x_2, \dots, x_1, \dots, x_n]^T$$

gdzie T - symbol transpozycji, jako

$$\|x\|_p = \left( \sum_{i=1}^n |x_i|^p \right)^{1/p}, \quad p \geq 1$$

gdzie  $|x_i|$  - moduł i-tej składowej wektora x.

Do konstrukcji praktycznych reguł ZPI skorzystamy z trzech norm szczególnych, wynikających z p-normy i posiadających interpretację geometryczną:

$$\|x\|_1 = \sum_{i=1}^n |x_i| \quad - \text{norma pierwsza (manhattanowska [7])}$$

$$\|x\|_2 = \left( \sum_{i=1}^n |x_i|^2 \right)^{1/2} = (x^T x)^{1/2} \quad - \text{norma euklidesowa}$$

$$\|x\|_\infty = \max_1 |x_i| \quad - \text{norma Czebyszewa (równomiernej zbieżności [7])}$$

Oznaczając przez  $\varepsilon_i$ ,  $i = \overline{1, n}$ , zadane przez użytkownika algorytmu optymalizacji, dopuszczalne tolerancje (błędy) określenia poszczególnych składowych wektora  $\hat{x}$ , gdzie  $\hat{x}$  jest estymatą (oceną) wektora  $x$ , można zaproponować kilka reguł ZPI. Wprowadzimy przy tym następujące oznaczenia:

$D = D(\hat{\Delta}x_k^i) \in \{S, K\}$  - decyzja (zmienna logiczna), będąca funkcją wyniku obliczeń  $\hat{\Delta}x_k^i$ , przyjmująca wartości logiczne ze zbioru dwuelementowego, w którym oznaczono:

$S$  - stop, tzn. zatrzymanie procesu obliczeń,

$K$  - kontynuacja, tzn. dalsza realizacja obliczeń.

$\hat{\Delta}x_k^i = \hat{x}_{k+1}^i - \hat{x}_k^i$  - różnica wektorów ocen  $\hat{x}$  wektora  $x$  w sąsiednich iteracjach (krokach);  $k = 0, 1, 2, \dots$

$\hat{x}_k^i = \left[ \frac{x_{1k}}{\varepsilon_1}, \frac{x_{2k}}{\varepsilon_2}, \dots, \frac{x_{1k}}{\varepsilon_1}, \dots, \frac{x_{nk}}{\varepsilon_n} \right]^T$  - wektor ocen  $\hat{x}$  w kroku

$k$ -tym, ważony odpowiednimi tolerancjami  $\varepsilon_i$  wg ogólnej zależności

$$\hat{x}_k^i = E^{-1/2} \hat{x}_k, \quad E = \text{diag}[\varepsilon_1^2, \varepsilon_2^2, \dots, \varepsilon_1^2, \dots, \varepsilon_n^2].$$

1. Reguła wykorzystująca ważoną normę pierwszą

$$D(\hat{\Delta}x_k^i) = \begin{cases} S, & \text{jeżeli } \|\hat{\Delta}x_k^i\|_1 \leq 1 \\ K, & \text{jeżeli } \|\hat{\Delta}x_k^i\|_1 > 1 \end{cases} \quad k=0, 1, 2, \dots \quad (1)$$

gdzie:  $\|\hat{\Delta}x_k^i\|_1 \stackrel{\text{df}}{=} \sum_{i=1}^n \Delta x_{ik}^i = \sum_{i=1}^n \left| \frac{\Delta x_{ik}}{\varepsilon_i} \right|$ ,  $\Delta x_{ik} = x_{i,k+1} - x_{ik}$

W przypadku szczególnym, gdy  $\varepsilon_i = \varepsilon$  dla  $i = \overline{1, n}$ , powyższa reguła może być przedstawiona w postaci

### Reguły zatrzymywania algorytmów

$$D(\hat{\Delta x}_k^1) = \begin{cases} S, & \text{jeżeli } \|\hat{\Delta x}_k^1\|_1 \leq \varepsilon \\ K, & \text{jeżeli } \|\hat{\Delta x}_k^1\|_1 > \varepsilon \end{cases} \quad k = 0, 1, 2, \dots \quad (1')$$

2. Reguła wykorzystująca ważoną normę euklidesową

$$D(\hat{\Delta x}_k^1) = \begin{cases} S, & \text{jeżeli } \|\Delta x_k^1\|_2^2 \leq 1 \\ K, & \text{jeżeli } \|\Delta x_k^1\|_2^2 > 1 \end{cases} \quad k = 0, 1, 2, \dots \quad (2)$$

$$\text{gdzie } \|\hat{\Delta x}_k^1\|_2^2 \stackrel{\text{df}}{=} (\hat{\Delta x}_k^1)^T \Delta x_k^1 = \sum_{i=1}^n (\Delta x_{ik}^1)^2 = \sum_{i=1}^n \left( \frac{\Delta x_{ik}^1}{\varepsilon_1} \right)^2$$

w przypadku szczególnym, gdy  $\varepsilon_1 = \varepsilon$  dla  $i = \overline{1, n}$ , reguła ta przybiera postać

$$D(\hat{\Delta x}_k^1) = \begin{cases} S, & \text{jeżeli } \|\hat{\Delta x}_k^1\|_2^2 \leq \varepsilon^2 \\ K, & \text{jeżeli } \|\hat{\Delta x}_k^1\|_2^2 > \varepsilon^2 \end{cases} \quad k = 0, 1, 2, \dots \quad (2')$$

W regułach (2) i (2') celowo zastosowano kwadraty norm  $\|\hat{\Delta x}_k^1\|_2$  i  $\|\Delta x_k^1\|_2$ , ponieważ oblicza się je bezpośrednio jako sumy kwadratów odpowiednich składowych, podczas gdy do obliczenia samych norm należałoby dodatkowo obliczyć pierwiastki kwadratowe.

3. Reguła wykorzystująca ważoną normę Czebyszewa

$$D(\hat{\Delta x}_k^1) = \begin{cases} S, & \text{jeżeli } \|\hat{\Delta x}_k^1\|_\infty \leq 1 \\ K, & \text{jeżeli } \|\hat{\Delta x}_k^1\|_\infty > 1 \end{cases} \quad k = 0, 1, 2, \dots \quad (3)$$

gdzie

$$\|\hat{\Delta x}_k^1\|_\infty \stackrel{\text{df}}{=} \max_i |\Delta x_{ik}^1| = \max_i \left| \frac{\Delta x_{ik}^1}{\varepsilon_1} \right| = |\Delta x_k^1(\max)| = \left| \frac{\Delta x_k^1(\max)}{\varepsilon_{\min}} \right|$$

przy czym  $\max_i |\Delta x_{ik}^1 / \varepsilon_1|$  otrzymujemy dla  $\min_i |\varepsilon_1| = \varepsilon_{\min}$

Reguły decyzyjne (1)-(3) obejmują trzy klasy funkcji decyzyjnych, które to funkcje można przedstawić w jednej ogólnej postaci

$$\|\hat{\Delta x}_k^1\|_p^t = 1, \quad \text{gdzie } t = \begin{cases} 1 & \text{dla } p \neq 2 \\ 2 & \text{dla } p = 2 \end{cases} \quad (4)$$

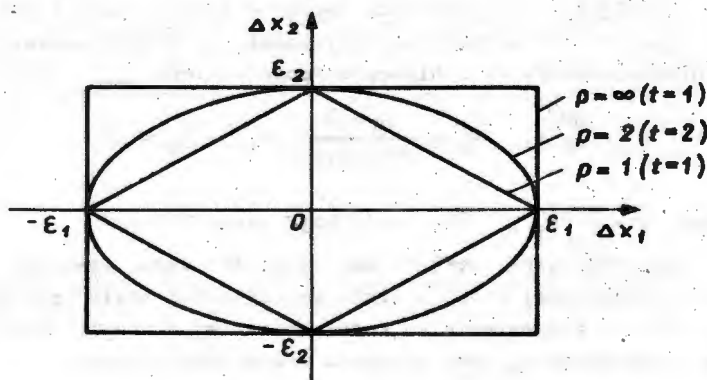
Funkcja (4) opisuje powierzchnię, rozgraniczającą n-wymiarową przestrzeń estymowanych parametrów  $x_1$  na dwa wzajemnie

dopełniające się obszary:

- 1) zamknięty, opisany nierównością  $\|\hat{\Delta x}_k^i\|_p^t \leq 1$  (obszar stopu),
- 2) otwarty, opisany nierównością  $\|\hat{\Delta x}_k^i\|_p^t > 1$  (obszar kontynuacji).

Dla  $p = 1$  obszar zamknięty jest  $n$ -wymiarowym wielościanem wypukłym. W przypadku szczególnym, dla jednakowych tolerancji  $\varepsilon_i = (i = \overline{1, n})$ , otrzymujemy  $n$ -wymiarowy odpowiednik ośmiościanu foremnego (oktaedru).

Dla  $p = 2$  obszar zamknięty jest hiperelipsoide; w przypadku jednakowych tolerancji dla wszystkich składowych otrzymujemy hiperkulę. Dla  $p = \infty$  obszar zamknięty przybiera postać hiperprostokąta; w przypadku jednakowych tolerancji otrzymujemy hipersześcian.



Rys.1. Kształt funkcji decyzyjnych  $\|\Delta x^i\|_p^t = 1$  dla  $n = 2$

Na rys.1 przedstawiono geometryczną interpretację funkcji decyzyjnych jako granic między obszarami stopu i kontynuacji dla przypadku dwuwymiarowego wektora parametrów  $x = [x_1, x_2]^T$ . Granice te, określone regułami decyzyjnymi (1), (2) i (3) tworzą kontury odpowiednio: rombu wpisanego w elipsę (użycie ważonej normy pierwszej), elipsy (ważona norma euklidesowa), prostokąta opisanego na elipsie (ważona norma Czebyszewa).

### Reguły zatrzymywania algorytmów

#### 3. Wyniki badań i wnioski końcowe

W przypadku ogólnym wielowymiarowe bryły, tworzące wspomniane obszary zamknięte, pozostają - względem siebie w następującym związku geometrycznym: hiperwielością jest wpisana w hiperelipsoidę, która z kolei jest wpisana w hiperprostopadłością. Wynika stąd, że reguła (1) jest, powiedzmy, najbardziej ostra spośród ww. reguł, tzn. przy jej użyciu zatrzymanie obliczeń może nastąpić po liczbie iteracji większej niż przy użyciu reguł pozostałych. Dochodzimy więc do zagadnienia tzn. ostrości reguły ZPI.

Mówimy, że dana reguła ZPI jest bardziej ostra od innej reguły ZPI, jeżeli zatrzymanie procesu iteracyjnego jest dla tej reguły trudniejsze niż dla reguły, z którą dana reguła jest porównywana.

Bezwzględną miarą ostrości przedstawionych reguł ZPI może być objętość odpowiednich obszarów zamkniętych. I tak, dla reguł (1)-(3) otrzymujemy objętości  $V_p$ : hiperwielościenu  $V_1$ , hiperelipsoidy  $V_2$  i hiperprostopadłością  $V_\infty$ , a więc

$$V_1 = \frac{2^n}{n!} V, \quad V_2 = \frac{\pi^{n/2}}{\Gamma(n/2+1)} V, \quad V_\infty = 2^n V$$

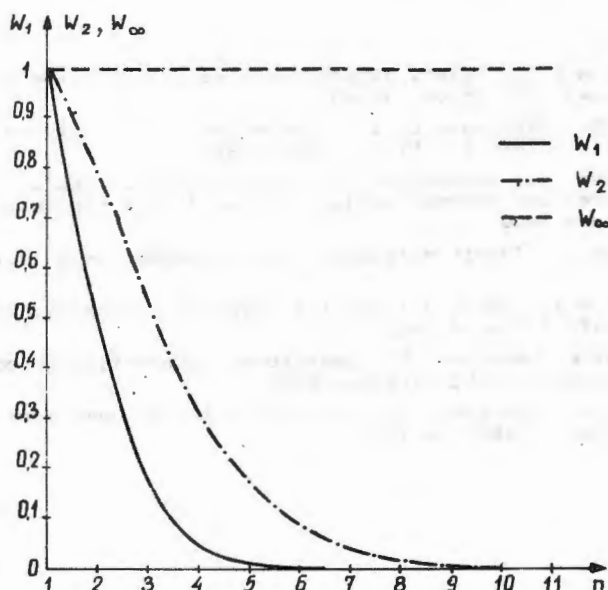
gdzie:  $V = \prod_{i=1}^n \varepsilon_i$ ,  $\Gamma(\cdot)$  - funkcja gamma Eulera.

Względną miarę ostrości ww. reguł ZPI można utworzyć z miary bezwzględnej w odniesieniu do objętości bryły największej (tzn. w odniesieniu do reguły najmniej ostrej). Uzyskujemy zatem miarę  $W_p$  jako objętość unormowaną, a więc

$$W_1 = V_1/V_\infty, \quad W_2 = V_2/V_\infty, \quad W_\infty = 1$$

Na podstawie ww. miary względnej  $W_p$  oraz dokonanych obliczeń  $W_p$  dla  $p = 1, 2, \infty$ , uzyskano wykresy (rys.2), z których można łatwo oszacować stopień trudności spełnienia warunku zatrzymania obliczeń przy użyciu reguł (1) i (2) w odniesieniu do reguły (3) jako najłatwiejszej do spełnienia (najłżejszej).





Rys.2.2. Zmiana objętości obszarów zatrzymania obliczeń przy użyciu reguł (1) i (2) w porównaniu z regułą (3), w funkcji wymiaru przestrzeni parametrów

Z rys.2 widać, że ze wzrostem liczby składowych wektora  $x$  następuje istotne pogorszenie możliwości szybkiego zatrzymania obliczeń, szczególnie dla reguły (1). Przykładowo, tylko dla  $n = 3$  spełnienie warunku zatrzymania obliczeń będzie dla reguły (1) 6-krotnie, a dla reguły (2)  $6/\pi \approx 2$ -krotnie trudniejsze niż dla reguły (3). Dla  $n = 6$  (np. trzy składowe wektora położenia i tyleż składowych wektora prędkości) pogorszenie będzie odpowiednio: 720-krotne dla reguły (1) i  $384/\pi^3 \approx 12$ -krotne dla reguły (2).

Zatem w przypadku optymalizacji wielkości wektorowych o dużej liczbie składowych  $n$  należy liczyć się z tym, że zastosowanie ostrych reguł (1) i (2) może prowadzić do długotrwałego procesu iteracyjnego, a nawet do niemożliwości poprawnego jego zakończenia (kumulacja błędów zaokrągleń i rozbieżność procesu). W takim przypadku lepiej zastosować regułę (3).

**ISBN 83-900412-1-9.**