Karuš, Kun-Takerovi uslovi

Zoran D. Jeličić
Mirna N. Kapetina
2. novembar 2022.

1 Uvodna razmatranja

Ovo poglavlje je takođe posvećeno problemima optimizacije uz ograničenja tipa nejednakosti, odnosno iznalaženju ekstrema kriterijuma optimalnosti

$$y = y(x_1, x_2, \dots, x_n) \ , \tag{1}$$

ako su promenljive stanja x_i međusobno ograničene skupom nejednakosti

$$g_k(x_1, x_2, \dots, x_n) \leq 0 \quad \text{za} \quad k = 1, 2, \dots, m \ . \tag{2}$$

Postupak Karuš, Kun-Takera (KKT) je prirodno uopštenje metoda Lagranževih množitelja u rešavanju optimizacionih problema sa ograničenjima tipa nejednakosti. Uopštenje će ići u dva pravca, prvi smanjivanje dimenzionalnosti prilikom rešavanja i drugi korišćenje znaka Langranževih množitelja da bi odredili karakter ekstrema. Podsećamo da smo ovu mogućnost, samo nagovestili u poslednjem primeru prethodnog poglavlja, a sada ćemo je staviti u mnogo formalniji okvir.

Ovaj postupak je u literaturi poznat i kao Kun-Takerovi uslovi, po Harold W. Kuhn i Albert W. Tucker, koji su ih publikovali 1951. godine. Nedavno se ispostavilo da je do skoro identičnog rezulata došao i William Karush u svojoj magistarskoj tezi 1939. godine, stoga i ovako složen naslov postupka. U literaturi, posebno onoj koja je prva izdanja imala pre 2000. godine, ostao je termin Kun-Takerovi uslovi, nezasluženo izostavljajući profesora Karuša sa California State University, Northridge. Važno je reći da je zbog svoje jednostavnosti ovaj postupak inherentno sadržan u velikom broju algoritama mašinskog učenja.

1.1 Karuš, Kun-Takerovi uslovi, teorijske osnove

Radi potpunosti, podsetićemo na metod Lagranževih množitelja. U slučaju da je kriterijum optimalnosti (1) ograničen nejednakostima (2), uz uvođenje dodatnih promenljivih (x_{n+k}),

$$\Phi_k = g_k(x_1, x_2, \dots, x_n) + x_{n+k}^2 = 0 \quad \text{za} \quad k = 1, 2, \dots, m \ , \tag{3}$$

Lagranževa funcija ima sledeći oblik

$$F = y(x_1, x_2, \dots, x_n) + \sum_{k=1}^{m} \lambda_k \left(g_k + x_{n+k}^2 \right) \ . \tag{4}$$

Parcijalni izvodi Lagranžijana po svim promenljivima daju potrebe uslove i to prvih n po originalnim x_i

$$\frac{\partial F}{\partial x_i} = \frac{\partial y(x_i)}{\partial x_i} + \sum_{k=1}^{m} \lambda_k \frac{\partial g_k(x_i)}{\partial x_i} = 0 \quad \text{za} \quad i = 1, 2, \dots, n \ , \tag{5}$$

a preostalih m po parazitskim varijablama x_{n+k}

$$\frac{\partial F}{\partial x_{n+k}} = 2\lambda_k x_{n+k} = 0 \quad \text{za} \quad k = 1, 2, \ldots, m \;. \tag{6}$$

Podsetićemo se na tri moguća rešenja izraza (6)

1. $\lambda_k = 0$ i $x_{n+k} \neq 0$
   Iz (3) logično sledi da je rešenje u oblasti $g_k = -x_{n+k} < 0$.

2. $\lambda_k \neq 0$ i $x_{n+k} = 0$
   Iz (3) sledi da je rešenje na granici $g_k = 0$.

3. $\lambda_k = 0$ i $x_{n+k} = 0$
   Iz (3) sledi da je rešenje na granici $g_k = 0$.

Posmatrajući ova tri moguća rešenja, nije teško primetiti sledeću zakonomernost. U slučaju da je dodatna promenljiva jednaka nuli $x_{n+k} = 0$, rešenje je na granici, odnosno važi da je $g_k = 0$. Ovo praktično znači, da je izraz (6) matematički ekvivalentan sa izrazom

$$\lambda_k g_k = 0 \quad \text{za} \quad k = 1, 2, \ldots, m \;. \tag{7}$$

Ako prihvatimo činjenicu da su izrazi (6) i (7) matematički ekvivalenti i ako znamo da dodatne promenljive x_{n+k} ne participiraju ni u jednim drugim potrebnim uslovima (5), otvara nam se mogućnost da ove parazitske promenljive x_{n+k} potpuno izostavimo i time redukujemo dimenzionalnost našeg problema. Napomena, ovo ne znači da ove promenljive ne postoje formalno (3), ali ih možemo izostaviti u rešavanju našeg optimizacionog problema tako što ćemo jednačinu $\lambda_k x_{n+k} = 0$, zameniti ekvivalentnom $\lambda_k g_k = 0$. Ovu smenu ćemo eksplicitno prikazati u algoritmu, koji ćemo uvesti na kraju ovog poglavlja.

U nastavku teksta pokušaćemo da uspostavimo relaciju između karaktera ekstrema i znaka Lagranževih množitelja, podsećamo da smo na ovu vezu, ali bez formalnih dokaza ukazali u primeru na kraju prethodnog poglavlja.

Počećemo našu studiju pod pretpostavkom da kriterijum optimalnosti $y(x_i)$ zavisi od n promenljivih stanja

$$y = y(x_1, x_2, \ldots, x_n) \;, \tag{8}$$

ali uz samo jedno ograničnje tipa nejednakosti [1]

$$g(x_1, x_2, \ldots, x_n) \leq 0 \;. \tag{9}$$

Pošto ispitujemo vezu između znaka λ i karaktera rešenja, **pretpostavili smo** da je u tački ekstrema $\mathbf{x}^*$, $\lambda > 0$. U tom slučaju, ako

[1] Ovu pretpostvaku smo uveli bez gubitka na opštosti, samo da bi pojednostavili matematičku proceduru i da bi lakše naglasili ključne korake, koji bi bili manje vidljivi u rigidnijem matematičkom postupku.

pogledamo jednačine (6) i (7) dodatna promenljiva x_{n+1} mora biti jednaka nuli, odnosno rešenje je na granici. Sada se Lagranžijan (4) u našem slučaju sa jednim ograničenjem i uz $x_{n+1} = 0$ zapisuje u sledećem obliku

$$F = y(x_1, x_2, \dots, x_n) + \lambda g(x_1, x_2, \dots, x_n) \ . \tag{10}$$

Iz potrebnih uslova ekstrema (5) znamo da u tački $\mathbf{x}^*$ važe sledeće relacije

$$\frac{\partial F}{\partial x_i} = \frac{\partial y}{\partial x_i} + \lambda \frac{\partial g}{\partial x_i} = 0 \quad \text{za} \quad i = 1, 2, \dots, n \ , \tag{11}$$

odnosno da u tački $\mathbf{x}^*$ važi

$$\frac{\partial y}{\partial x_i} = -\lambda \frac{\partial g}{\partial x_i} \quad \text{za} \quad i = 1, 2, \dots, n \ . \tag{12}$$

Prethodni izraz (12) možemo zapisati i na sledeći način

$$\sum_{i=1}^{n} \frac{\partial y}{\partial x_i} dx_i = -\lambda \sum_{i=1}^{n} \frac{\partial g}{\partial x_i} dx_i \ , \tag{13}$$

ili kao

$$dy = -\lambda dg \ . \tag{14}$$

Gde su dy i dg priraštaji funkcija y i g respektivno, za dopustiva pomeranja dx_i. Mi tvrdimo da će za pretpostavljenu vrednost Lagranževog množitelja $\lambda > 0$ rešenje biti u minimumu. Ovo ćemo pokušati i grafički da ilustrujemo, slika 1, a iz očiglednih razloga tu studiju ćemo ograničiti na samo dve promenljive stanja x_1 i x_2.

Na osnovu analize slike 1, nije teško zaključiti da su iz tačke ekstrema $\mathbf{x}^*$ dozvoljeni samo priraštaji $dg \leq 0$. Ako ovu vrednost unesemo u izraz (14) uz pretpostavku da je $\lambda > 0$, lako izračunavamo da je tada priraštaj $dy \geq 0$, odnosno da iz tačke ekstrema vrednost funkcije y ne opada ili da je u tačaki $\mathbf{x}^*$ minimum funkcije (8).

Razmatranja iz ovog poglavlja, pokušaćemo da formalizujemo u algoritamu za rešavanje optimizacionih problema sa ograničenjima tipa nejednakosti uz oslonac na metod Karuš, Kun-Takera.

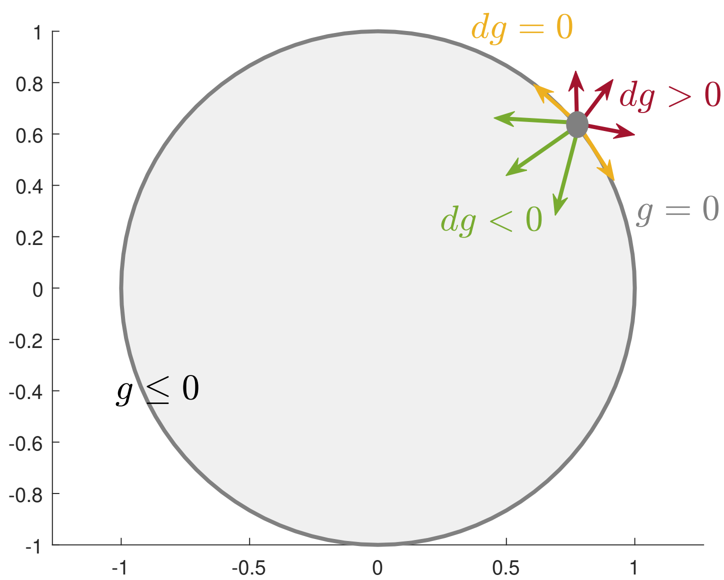
Formalizam Karuš, Kun-Takera

Kriterijum optimalnosti, zavisi od n promenljivih stanja

$$y = y(x_1, x_2, \dots, x_n) \ , \tag{15}$$

gde su promenljive stanja x_i međusobno ograničene skupom nejednakosti

$$g_k(x_1, x_2, \dots, x_n) \leq 0 \quad \text{za} \quad k = 1, 2, \dots, m \ . \tag{16}$$



Slika 1: Na grafiku je kružnicom predstavljeno ograničenje $g \leq 0$, dozvoljene vrednosti su unutar kružnice $g < 0$ i na samoj kružnici $g = 0$. Pretpostavkom da je $\lambda > 0$ naterali smo da eksterm $\mathbf{x}^*$ bude na samoj kružnici (siva tačka na crtežu). Iz te tačke dozvoljena su pomeranja u ununutrašnjost kružnice $dg < 0$ tj. $g + dg < 0$ (zelene strelice), kao i da ostanemo na samoj kružnici $dg = 0$ tj $g + dg = 0$ (žute strelice). Nisu dozvoljena pomeranja van kružnice $dg > 0$ tj. $g + dg > 0$ (bordo strelice). Zaključak je da su iz tačke ekstrema na granici dozvoljeni svi priraštaji dg, koji zadovoljavaju uslov $dg \leq 0$.

Pod uobičajenom pretpostavkom o diferencijabilnosti razmatranih funkcija, formalizam možemo predstaviti na sledeći način.

1. Formirati prošireni kriterijum optimalnosti F u formi Lagranžijana, **bez dodatnih varijabli** x_{n+k}

$$F = y(x_1, x_2, \ldots, x_n) + \sum_{k=1}^{m} \lambda_k g_k(x_1, x_2, \ldots, x_n) \, . \tag{17}$$

   Kao što smo u uvodu ovog poglavlja pokazali, dodatne promenljive postoje samo u izrazima (6) i one se uspešno mogu zameniti ekvivalentnim izrazom (7). Uvođenje dodatnih promenljivih u ovom koraku, bila bi formalna greška.

2. Parcijalni izvodi po svim originalnim promenljivima moraju biti jednaki nuli

$$\frac{\partial F}{\partial x_i} = \frac{\partial y(x_i)}{\partial x_i} + \sum_{k=1}^{m} \lambda_k \frac{\partial g_k(x_i)}{\partial x_i} = 0 \quad \text{za} \quad i = 1, 2, \ldots, n \, . \tag{18}$$

3. Sledeća jednačina, predstavlja zamenu izraza (6) po principima, koje smo ranije usvojili

$$\lambda_k g_k = 0 \quad \text{za} \quad k = 1, 2, \ldots, m \, . \tag{19}$$

4. Sledeći korak bi bio rešiti sisteme jednačina (18)-(19) i proveriti da li su zadovoljena ograničenja (16).

5. Na osnovu izračunatih vrednosti Lagranževih množitelja, diskutovaćemo karakter ekstrema po već utvrđenim principima

   (a) Akko su vrednosti $\lambda_k \geq 0$, dobijena tačka je **minimum**
   (b) Akko su vrednosti $\lambda_k \leq 0$, dobijena tačka je **maksimum**
   (c) U svim ostalim slučajevima, radi se o prevojnoj tački.

   Napomena uslov $\lambda_k \geq 0$ znači da je bar jedna vrednost Lagranževih množitelja različita od nule i veća od nule. Slično je i u slučaju maksimuma.

Izloženi algoritam ilustrovaćemo na primeru.

Primer 1. Ilustracija metoda, Karuš, Kun-Takera

Metodom Karuš, Kun-Takera, želimo da nađemo ekstreme funkcije i ispitamo njihov karakter

$$y(x_1, x_2) = x_1^3 + 2x_2^2 - 10x_1 + 6 + 2x_2^3 \,.$$

Uz ograničenja

$$\begin{aligned} x_1 x_2 &\leq 10 \\ x_1 &\geq 0 \\ x_2 &\geq 0 \,. \end{aligned}$$

ili u formi, koja nam je potrebna

$$g_1 = x_1 x_2 - 10 \leq 0 \quad g_2 = -x_1 \leq 0 \quad g_3 = -x_2 \leq 0$$

Nastavljamo formiranjem Lagranžijana, uz važnu napomenu, da se **sva** ograničenja moraju uvrstiti u Lagranžijan [2], bez obzira koliko su trivijalna.

[2] Karakter ekstrema zavisi od svih ograničenja odnosno od odgovarajućih Lagranževih množitelja, koji ih uvode u prošireni kriterijum optimalnosti, što će biti jasno već u ovom primeru.

$$F = x_1^3 + 2x_2^2 - 10x_1 + 6 + 2x_2^3 + \lambda_1 (x_1 x_2 - 10) + \lambda_2 (-x_1) + \lambda_3 (-x_2) \,.$$

Potrebni uslovi ekstrema se dobijaju parcijalnim izvodom Lagranževe funkcije po originalnim promenljivima

$$\begin{aligned} \frac{\partial F}{\partial x_1} &= 3x_1^2 - 10 + \lambda_1 x_2 - \lambda_2 = 0 \\ \frac{\partial F}{\partial x_2} &= 4x_2 + 6x_2^2 + \lambda_1 x_1 - \lambda_3 = 0 \,, \end{aligned}$$

uz uslove

$$\begin{aligned} \lambda_1 (x_1 x_2 - 10) &= 0 \\ \lambda_2 (-x_1) &= 0 \\ \lambda_3 (-x_2) &= 0 \,. \end{aligned}$$

U rešavanju ovog sistema jednačina, treba voditi računa o rešenjima, koja se međusobno isključuju, što može znatno olakšati studiju problema. Tako na primer ako je $\lambda_1 \neq 0$, odnosno $x_1 x_2 = 10$, isključuje mogućnost $x_1 = x_2 = 0$, već se nameće $\lambda_2 = 0$, odnosno $\lambda_3 = 0$. Pažljivim rešavanjem dobijamo sledeće stacionarne tačke.

	A	B	C	D	E
x_1	3.847	−3.568	0	1.8257	−1.8257
x_2	2.6	−2.802	0	0	0
λ_1	−13.238	10.063	0	0	0
λ_2	0	0	−10	0	0
λ_3	0	0	0	0	0
g_1	0	0	−10	−10	−10
g_2	−3.847	3.568	0	−1.8257	1.8257
g_2	−2.6	2.802	0	0	0
karakter	Maksimum	$g_k > 0$	Maksimum	Ne znamo	$g_k > 0$

Tabela 1: Rešenje problema metodom Karuš, Kun-Takera. Tačke B i E ne zadovoljavaju ograničenja, tako da nisu validna rešenja. Tačka D ima sve Lagranževa množitelje jednake nuli, odnosno metodom KKT ne možemo suditi o karakteru rešenja. Međutim, karakter ovog rešenja je minimum, razmislite kako ovo možemo formalno proveriti. Tačke A i C su na osnovu znaka λ_k maksimum.

U nastavku vam dajemo dva primera za samostalan rad, uzimajući u obzir značaj ovog postupka i učestanost pojavljivanja na ispitu, savetujemo da ih pažljivo rešite.

Primer 2. KKT jedno ograničenje, parametri a i b

Naći minimum funkcije

$$y = 4\,(x_1 - a)^2 + 5\,(x_2 - b)^2 \ ,$$

uz ograničenje

$$x_1 + x_2 \geq 1 \ .$$

Rešenje diskutovati u zavinsoti od međusobnog odnosa parametara a i b, odnosno $(a + b)$.

Naredni primer će imati kratak teorijski uvod, a primer zaista zaslužuje posebnu pažnju

Primer 3. Metod KKT, kombinovana ograničenja

Cilj nam je da nađemo optimalne vrednosti kriterijuma optimalnosti

$$y = y(x_1, x_2, \ldots, x_n) \ ,$$

ako postoje dve vrste ograničenja

$$\begin{aligned} g_k(x_i) \leq& 0 \quad \text{za} \quad i = 1, 2, \ldots n. \quad k = 1, 2, \ldots m \\ h_j(x_i) =& 0 \quad \text{za} \quad i = 1, 2, \ldots n. \quad j = 1, 2, \ldots l \end{aligned}$$

Prvi korak je formiranje proširenog kriterijuma optimalnosti, gde će ograničenja tipa nejednakosti biti uvedena uz pomoć množitelja λ_k po pravilima algortima KKT, a za ograničenja tipa jednakosti koristićemo novi skup množitelja μ_j u konvencijalnom formalizmu Lagranževih množitelja za ovu klasu ograničenja.

$$F = y\left(x_i\right) + \sum_{k=1}^{m} \lambda_k g_k\left(x_i\right) + \sum_{j=1}^{l} \mu_j h_j\left(x_i\right) \quad \text{za} \quad i = 1, 2, \ldots n\ .$$

Dalje rešavanje se odvija u duhu obe teorije KKT i Lagranževih množitelja, ali strogo vodeći računa da se ovi postupci međusobne ne pomešaju. Nije teško zaključiti da se rešavanje ovog problema sprovodi po sledećoj proceduri

$$\begin{aligned} &\frac{\partial F}{\partial x_i} = \frac{\partial y}{\partial x_i} + \sum_{k=1}^{m} \lambda_k \frac{\partial g_k\left(x_i\right)}{\partial x_i} + \sum_{k=1}^{m} \mu_j \frac{\partial h_j\left(x_i\right)}{\partial x_i} = 0 \quad i = 1, 2, \ldots n \\ &g_k\left(x_i\right) \leq 0 \quad \text{za} \quad i = 1, 2, \ldots n. \quad k = 1, 2, \ldots m \\ &h_j\left(x_i\right) = 0 \quad \text{za} \quad i = 1, 2, \ldots n. \quad j = 1, 2, \ldots l \\ &\lambda_k g_k\left(x_i\right) = 0 \quad \text{za} \quad i = 1, 2, \ldots n. \quad k = 1, 2, \ldots m \\ &\lambda_k \geq 0 \quad \text{tačka je minimum} \\ &\lambda_k \leq 0 \quad \text{tačka je maksimum} \end{aligned}$$

Čitaocima ostavljamo da, uz razumevanje, ovaj način rešavanja primene na sledeći zadatak

Naći minimum funkcije

$$y = (x_1 - 1)^2 + x_2 - 2\ ,$$

uz ograničenja

$$\begin{aligned} h(x) =& x_2 - x_1 - 1 = 0 \\ g(x) =& x_1 + x_2 - 2 \leq 0\ . \end{aligned}$$