

Prilagodba podataka

Marko Imbrišak

21. rujna 2017.

Sažetak

U ovom je radu opisana prilagodba podataka primjenjujući metode diferencijalne geometrije. Također, opisana je metoda kako doći do Fisherove informacijske metrike na prostoru parametara. Navedeno je kako prilagoditi podatke pomoću dva različita načina. Prvi je način korištenje Levenberg-Marquardt metode koju se može generalizirati, a drugi pristup jest aproksimacijska metoda mnogostrukosti s rubom.

1 Uvod

Kompleksni modeli koji se koriste u različitim granama znanosti za opis složenih problema uključuju određeni broj slobodnih parametara koje moramo prilagoditi empirijskim podacima. Pritom često nemamo jasnu predodžbu koji su parametri nužni za precizan opis problema. Također, metode prilagodbe podataka imaju problema s modelima koji u prostoru podataka razapinju mnogostrukost s rubom.

U poglavlju *Fisherova informacijska metrika* (FIM) izvodimo metriku na prostoru parametara, u poglavlju *Prilagodba podataka* pomoću nje opisujemo prilagodbu podataka, u poglavlju *Sloppy problemi* definiramo općeniti statistički problem modela čija Fisherova informacijska metrika ima svojstvene vrijednosti koje se međusobno razlikuju za mnogo redova

veličine, u poglavlju *Geodezijska jednadžba* opisujemo metodu nalaženja geodezika za statističke probleme te ju u poglavlju *Primjena na sloppy probleme* primjenjujemo.

2 Fisherova informacijska metrika

Za model opisan parametrima p , mjerenja, $\{x_1, \dots, x_{N_d}\}$, su opisana distribucijom vjerojatnosti mjerenja uz dane parametre p , $p(x_1, \dots, x_{N_d}|p)$. Ako su mjerenja nezavisna, nju možemo izraziti kao produkt vjerojatnosti svakog pojedinog x_i za dani model opisan parametrima p

$$p(x_1, \dots, x_i, \dots, x_{N_d}|p) = \prod_{i=1}^{N_d} p(x_i|p). \quad (1)$$

Budući daje ovaj izraz dan produktom, korisnije je gledati log-vjerojatnost, $\ln p(x|p)$ jer je ona aditivna po x_i

$$p(x_1, \dots, x_i, \dots, x_{N_d}|p) = e^{\sum_{i=1}^{N_d} \ln p(x_i|p)}. \quad (2)$$

Negativno očekivanje log-vjerojatnosti zovemo *Shannonovom entropijom*,

$$S(p) = - \sum_{i=1}^{N_d} p(x_i|p) \ln p(x_i|p), \quad (3)$$

a ako promatramo dva modela, p_1 i p_2 , možemo promatrati očekivanje log-vjerojatnosti modela 2 za model 1, odnosno možemo promatrati međuentropiju

$$S(p_1||p_2) = - \sum_{i=1}^{N_d} p(x_i|p_1) \ln p(x_i|p_2). \quad (4)$$

Razliku međuentropija modela 1 i 2 zovemo *Kullback-Leiblerovom divergencijom*, $D_{KL}(p_1||p_2) = S(p_1||p_2) - S(p_1)$, odnosno

$$D_{KL}(p_1||p_2) = \sum_{i=1}^{N_d} p(x_i|p_1) \ln \frac{p(x_i|p_1)}{p(x_i|p_2)}. \quad (5)$$

Promotrimo sada njezina svojstva. Za svaki $x > 0$ vrijedi $\ln x \leq (x - 1)$, stoga vrijedi

$$D_{KL}(p_1||p_2) = - \sum_{i=1}^{N_d} p(x_i|p_1) \ln \frac{p(x_i|p_2)}{p(x_i|p_1)} \quad (6)$$

$$\geq - \sum_{i=1}^{N_d} p(x_i|p_1) \left(\frac{p(x_i|p_2)}{p(x_i|p_1)} - 1 \right) \quad (7)$$

$$= \sum_{i=1}^{N_d} p(x_i|p_2) - p(x_i|p_1) \quad (8)$$

$$= -(1 - 1) = 0, \quad (9)$$

odnosno K-L divergencija je nenegativna, $D_{KL}(p_1||p_2) \geq 0$. Ovo svojstvo nam govori

kako bi K-L divergencija mogla poslužiti kao metrika u prostoru parametara. Za $p_1 = p_2$ vrijedi $D_{KL}(p_1||p_1) = 0$, a također vrijedi i obrat ove tvrdnje, $D_{KL}(p_1||p_2) = 0 \Rightarrow \ln p(x_i|p_1) - \ln p(x_i|p_2) = 0 \Rightarrow p(x_i|p_1) = p(x_i|p_2) \Rightarrow p_1 = p_2$. Međutim, K-L divergencija ne može biti metrika jer nije simetrična, što se vidi iz razlike $|D_{KL}(p_1||p_2) - D_{KL}(p_2||p_1)|$,

$$|D_{KL}(p_1||p_2) - D_{KL}(p_2||p_1)| = \quad (10)$$

$$= \left| \sum_{i=1}^{N_d} (p(x_i|p_1) + p(x_i|p_2)) \ln \frac{p(x_i|p_1)}{p(x_i|p_2)} \right| \quad (11)$$

$$> \left| \sum_{i=1}^{N_d} (p(x_i|p_1) + p(x_i|p_2)) \left(\frac{p(x_i|p_2)}{p(x_i|p_1)} - 1 \right) \right| \quad (12)$$

$$= \left| 1 - \sum_{i=1}^{N_d} \frac{p(x_i|p_2)^2}{p(x_i|p_1)} \right|. \quad (13)$$

Budući da ovaj izraz nije uvijek jednak nuli, K-L divergencija nije simetrična. Također, ona ne zadovoljava niti nejednakost trokuta, što vidimo iz sljedećeg razmatranja

$$D_{KL}(p_1||p_2) + D_{KL}(p_2||p_3) - D_{KL}(p_1||p_3) \quad (14)$$

$$= S(p_1||p_2) - S(p_1) + S(p_2||p_3) \quad (15)$$

$$- S(p_2) - S(p_1||p_3) + S(p_1) \quad (16)$$

$$= S(p_1||p_2) + S(p_2||p_3) - S(p_1||p_3) - S(p_2) \quad (17)$$

Općenito, ova razlika entropija nije nula, što vidimo uzimajući slučaj samo jednog mjerenja s $p(x_1|p_1) < p(x_1|p_2) < p(x_1|p_3)$. Tada vrijedi $S(p_1||p_2) < S(p_1)$ i $S(p_2||p_3) < S(p_2)$. U tom slučaju vrijedi

$$D_{KL}(p_1||p_2) + D_{KL}(p_2||p_3) - D_{KL}(p_1||p_3) \quad (18)$$

$$< S(p_1) + S(p_2) - S(p_1||p_3) - S(p_2) \quad (19)$$

$$= -D_{KL}(p_1||p_3) \quad (20)$$

$$< 0. \quad (21)$$

Dakle, ne vrijedi uvijek izraz $D_{KL}(p_1||p_2) + D_{KL}(p_2||p_3) \geq D_{KL}(p_1||p_3)$. Iz ovoga vidimo da KL-divergencija nije dobro definirana metrika jer, iako je nenegativna i jednaka je nuli akko su dva modela ista, nije simetrična na zamjenu modela te ne zadovoljava nejednakost trokuta.

S druge strane, možemo pogledati kako se ponašaju mala odstupanja $p_2^\alpha = p_1^\alpha + \delta^\alpha$ od minimuma vjerojatnosti, p_1 ,

$$D_{KL}(p_1||p_2) = D_{KL}(p_1||p_1) + D_{KL,\alpha}\delta^\alpha \quad (22)$$

$$+ \frac{1}{2}D_{KL,\alpha\beta}\delta^\alpha\delta^\beta + O(\delta^3), \quad (23)$$

gdje smo koristili pokratu $\partial/\partial p^\alpha =_{,\alpha}$. Prvi član, $D_{KL}(p_1||p_1)$, jednak je nuli, a drugi član je očekivanje derivacije log-vjerojatnosti

$$D_{KL,\alpha} = \sum_{i=1}^{N_d} p(x_i|p_1)(\ln p(x_i|p_2))_{,\alpha} = E[(\ln p)_{,\alpha}]. \quad (24)$$

Budući da je minimum u p_1 , i očekivanje ove logaritamske derivacije je nula

$$E[(\ln p)_{,\alpha}] = 0. \quad (25)$$

Ostao je samo član

$$g_{\alpha\beta} = -\frac{1}{2} \sum_{i=1}^{N_d} p(x_i|p_1)(\ln p(x_i|p_2))_{,\alpha\beta}, \quad (26)$$

kojeg zovemo *Fisherovom informacijskom metrikom* (nadalje, FIM). Pišemo

$$\partial_{\alpha\beta} \ln p = \partial_\alpha(p\partial_\beta \ln p) - p\partial_\alpha \ln p \partial_\beta \ln p. \quad (27)$$

Koristeći jednadžbu (25), dolazimo do ekvivalentnog

izraza

$$g_{\alpha\beta} = \frac{1}{2} \sum_{i=1}^{N_d} p(x_i|p_1)(\ln p(x_i|p_2))_{,\alpha}(\ln p(x_i|p_2))_{,\beta}. \quad (28)$$

Ako s \mathbf{J} označimo logaritamski Jakobijan, $\mathbf{J} = \nabla \ln p(x_i|p)$, vidimo da je metrički tenzor očekivana vrijednost

$$g_{\alpha\beta} = \frac{1}{2} E[J_\alpha J_\beta]. \quad (29)$$

Budući da je ovaj izraz simetričan na zamjenu indeksa, $g_{\alpha\beta} = g_{\beta\alpha}$, za male pomake δ , KL-divergencija jest metrika u prostoru parametara.

3 Prilagodba podataka

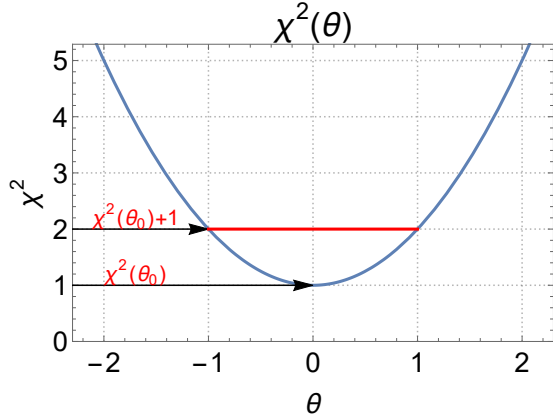
Za dati model koji ima N_p parametara promatra se N_p -dimenzionalni vektorski prostor pri čemu svaki parametar p^α predstavlja jednu dimenziju tog vektorskog prostora. Taj prostor općenito nije ravan, već je snabdjeven metrikom koja uključuje i parametre modela i rezultate korištenih mjerenja.

Promotrimo sada skup od N_d mjerenja opservable \mathcal{O} sa gdje su izmjerene vrijednosti označene s $\{\mathcal{O}_i^{exp}\}_{i \in \{1, \dots, N_d\}}$. Pretpostavljamo da je prava vrijednost opservable u i -tom mjerenju \mathcal{O}_i samo pomaknuta za gausijanski distribuiran šum, ξ_i od \mathcal{O}_i^{exp}

$$\mathcal{O}_i^{exp} = \mathcal{O}_i + \xi_i, \quad (30)$$

gdje šum ξ_i predstavljamo slučajnom varijablom ξ distribucije

$$p(\xi) \propto e^{-\xi^2/2}. \quad (31)$$



Slika 1: Shema ovisnosti funkcije $\chi^2(\theta)$ o jednodimenzionalnom parametru θ . Točka najbolje prilagodbe je označena s θ_0 . Crvenom je linijom prikazan interval s prihvatljivom vrijednosti parametra θ , $\chi^2(\theta_0) \leq \chi^2(\theta) \leq \chi^2(\theta_0) + 1$. Slika prilagođena iz [1].

Razliku eksperimentalno određene vrijednosti opservable i teorijskog modela, $\mathcal{O}_i^{exp} - \mathcal{O}_i$, zovemo *rezidual*.

Nadalje, pomoću reziduala možemo konstruirati funkciju procjene kvalitete modela:

$$\chi^2 = \sum_i^{N_d} \frac{(\mathcal{O}_i(p) - \mathcal{O}_i^{exp})^2}{\Delta \mathcal{O}_i^2}, \quad (32)$$

pri čemu je \mathcal{O}_i teorijski predviđena vrijednost, \mathcal{O}_i^{exp} eksperimentalno određena vrijednost, a $\Delta \mathcal{O}_i^2 = (\Delta \mathcal{O}_i^{exp})^2 + (\Delta \mathcal{O}_i^{num})^2 + (\Delta \mathcal{O}_i^{the})^2$ je rezultatna nepouzdanost uzrokovana eksperimentalnom, numeričkom i teorijskom nepouzdanosti. Kako bi prostoru parametara pridijelili metriku, koja je simetrični tenzor reda dva, koristimo Taylorov razvoj χ^2 oko njezinog minimuma, odnosno točke najbolje prilagodbe. Okolina u kojoj funkcija procjene ima prihvatljivu vrijednost određuje se uvjetom

$$\chi^2(p) \leq \chi^2(p_0) + 1. \quad (33)$$

Potom razvijemo χ^2 do drugog reda, koristeći Einsteinovu konvenciju o sumaciji:

$$\chi^2(p) - \chi^2(p_0) = \partial_\alpha \chi^2(p_0)(p^\alpha - p_0^\alpha) \quad (34)$$

$$+ \frac{1}{2} \partial_{\alpha\beta} \chi^2(p_0)(p^\alpha - p_0^\alpha)(p^\beta - p_0^\beta) \quad (35)$$

$$+ O(p^3), \quad (36)$$

iz čega dobivamo Hessijan od χ u točki najbolje prilagodbe

$$g_{\alpha\beta} = \frac{1}{2} \partial_{\alpha\beta} \chi^2(p_0). \quad (37)$$

Metriku možemo pojednostaviti ako χ^2 zapišemo pomoću reziduala

$$f_i(p) = \frac{\mathcal{O}_i(p) - \mathcal{O}_i^{exp}}{\Delta \mathcal{O}_i}. \quad (38)$$

U tom je slučaju metrika dana s

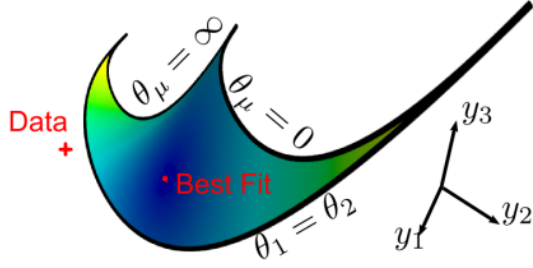
$$g_{\alpha\beta} = \sum_i^{N_d} (\partial_\alpha f_i \partial_\beta f_i + f_i \partial_{\alpha\beta} f_i), \quad (39)$$

gdje su parcijalne derivacije derivacije po parametrima, $\partial_\alpha = \frac{\partial}{\partial p^\alpha}$. Uzimajući u obzir da f_i izvrjednjujemo u p_0 , dobivamo izraz

$$g_{\alpha\beta} = \sum_i^{N_d} \partial_\alpha f_i \partial_\beta f_i. \quad (40)$$

Vidimo da smo dobili simetrični tenzor reda 2 koji je određen mjerenjima i teorijskim predviđanjima opservabli. Pomoću nje možemo odrediti udaljenost između dviju infinitezimalno bliskih točaka u prostoru parametara

$$ds^2 = g_{\alpha\beta} dp^\alpha dp^\beta, \quad (41)$$



Slika 2: Model s dvije eksponencijale izvrjednjen u točkama $t = 1/3, 1, 3$ za različite kombinacije parametara $\theta_{1,2}$. Slika preuzeta iz članka [2].

iz čega vidimo da prostor parametara možemo snabdjeti Fisherovom metrikom.

4 Sloppy problemi

Promotrimo problem s dva pozitivna parametra, $f(t) = e^{-\theta_1 t} + e^{-\theta_2 t}$ izvrjednjen u točkama $\{t_i\}_{i \in \{1, \dots, N_d\}}$,

$$\mathbf{f} = f_i = f(t_i). \quad (42)$$

Vidimo da možemo promatrati ovaj model izvrjednjen u točkama kao višedimenzionalni prostor. Ako mjerenja opisujemo točkama \mathbf{y} , χ^2 možemo zapisati kao

$$\chi^2(\theta^\mu) = (\mathbf{y} - \mathbf{f}(\theta^\mu))^T (\mathbf{y} - \mathbf{f}(\theta^\mu)). \quad (43)$$

Sa slike 2 vidimo da model obuhvaća samo dio cijelog prostora dimenzije N_d . Naime, u slučaju $\theta_1 = 0$, model daje $f(t|0, \theta_2) = 1 + e^{-\theta_2 t}$. Svaki $\theta_1 > 0$ će dati manji f , odnosno $f(t, \infty, \theta_2) < f(t, \theta_1, \theta_2) < f(t, 0, \theta_2)$. Usporedimo također slučaj

$\theta_1 = \theta_2$, $f(t, \theta_2, \theta_2) = 2f(t, \infty, \theta_2)$. Ovim različitim kombinacijama varijabli, omeđili smo se na samo dio prostora, odnosno mnogostrukost posjeduje rub.

Klasična metoda minimizacije je Levenberg-Marquardt metoda koja se temelji na minimizaciji χ^2 funkcije. U svakom koraku ona bira korak, δ^α u prostoru parametara. Korak se izvodi pomoću razvoja funkcije f ,

$$\mathbf{f} = \mathbf{f}_0 + \mathbf{J}_\alpha \delta^\alpha, \quad (44)$$

gdje je \mathbf{J}_α Jakobijan, definiran po komponentama, $\mathbf{J}_\alpha = J_{i\alpha} = \partial_\alpha f(t = t_i)$. S ovim pomakom χ^2 postaje

$$\chi^2 = \mathbf{r}^T \mathbf{r} - 2\mathbf{J}_\alpha^T \delta^\alpha \mathbf{r} + \mathbf{J}_\alpha^T \delta^\alpha \mathbf{J}_\beta \delta^\beta, \quad (45)$$

gdje smo koristili pokratu za rezidual, $\mathbf{r} = \mathbf{y} - \mathbf{f}_0$. Minimiziranjem izraza (45) po δ^α dobivamo jednadžbu

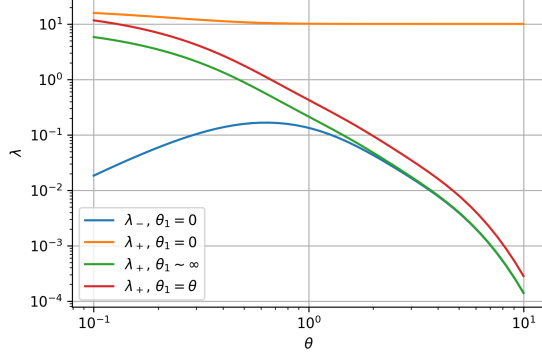
$$\mathbf{J}_\mu^T \mathbf{J}_\nu \delta^\nu = \mathbf{J}_\mu^T \mathbf{r}. \quad (46)$$

U ovom izrazu možemo prepoznati FIM, $g_{\mu\nu} = \mathbf{J}_\mu^T \mathbf{J}_\nu$, a množenjem s $g^{\alpha\mu}$ dobivamo

$$\delta^\alpha = g^{\alpha\mu} \mathbf{J}_\mu^T \mathbf{r}. \quad (47)$$

Problem koji postoji u ovom slučaju je da postoji veliki raspon u veličini svojstvenih vrijednosti metrike, što se može vidjeti i iz jednostavnog eksponencijalnog modela. Uzmimo tri vrijednosti parametra t , $t = \{1/3, 1, 3\}$, tada je metrika u prostoru parametara razapetom s $\theta_{1,2}$ dana s

$$g(\theta_1, \theta_2) = \begin{pmatrix} \sum_{i=1}^3 t_i^2 e^{-2\theta_1 t_i} & \sum_{i=1}^3 t_i^2 e^{-(\theta_1 + \theta_2) t_i} \\ \sum_{i=1}^3 t_i^2 e^{-(\theta_1 + \theta_2) t_i} & \sum_{i=1}^3 t_i^2 e^{-2\theta_2 t_i} \end{pmatrix}. \quad (48)$$



Slika 3: Svojstvene vrijednosti metrike za slučajeve $\theta_1 = 0, \infty, \theta_2$.

Svojstvene vrijednosti ove metrike prikazane su na slici 3. Vidimo da se dvije svojstvene vrijednosti, λ_{\pm} , za određene vrijednosti parametara međusobno razlikuju i za više redova veličine.

To znači da će inverz metrike imati jako velike vrijednosti, t.j. biti numerički nestabilan. Levenberg-Marquardt metoda ovaj problem rješava izabirući najmanju vrijednost svojstvene vrijednosti za svaki pojedini korak, λ^* i definirajući novu metriku, \tilde{g} ,

$$\tilde{g}_{\mu\nu} = g_{\mu\nu} + \lambda^* \delta_{\mu\nu}. \quad (49)$$

Vidimo da ova promjena samo globalno pomiče svojstvene vrijednosti jer svojstvene vrijednosti tada računamo iz jednadžbe

$$\det(\tilde{g} - \lambda I) = \det(g - (\lambda - \lambda^*)I). \quad (50)$$

Nedostatak ove metode je da ne uspijeva uvijek pronaći zadovoljavajuće rješenje, stoga je korisno generalizirati ovu metodu kako bi se uključilo gibanje

po geodeziku, što je opisano u sljedećem poglavlju.

5 Geodezijska jednadžba

Uočavamo da će za *sloppy* problem distribucija svojstvenih vrijednosti Fisherove metrike biti eksponencijalna. Odredi se najmanja svojstvena vrijednost jer ona ima najveću nepouzdanost kako bi se odredio geodezik u smjeru njezinog svojstvenog vektora. Geodezik određujemo rješavajući geodezijsku jednadžbu

$$\dot{\theta}^\alpha \nabla_\alpha \dot{\theta}^\beta = 0, \quad (51)$$

gdje je ∇_α kovarijantna derivacija, a točka označava derivaciju po parametru τ kojim opisujemo geodezijsku krivulju. Treba napomenuti kako se τ ne odnosi se na vremensku ovisnost, koju fizikalni model može imati. Jednadžba (51) nam govori o tome da je tangentni vektor $\dot{\theta}^\beta$ konstantno duž tangentnog vektora $\dot{\theta}^\alpha$, odnosno da je tangentni vektor transportiran paralelno sebi samom [3]. Kako bismo napravili paralelni transport u zakrivljenom prostoru, potrebni su nam Christoffelovi simboli, objekti po formi slični tenzoru sačinjeni od prvih derivacija metriке. Gibanje po geodeziku opisujemo geodezijskom jednadžbom (51), a nju možemo transformirati u pogodan oblik raspisujući kovarijantnu derivaciju

$$\dot{\theta}^\alpha (\partial_\alpha \dot{\theta}^\beta + \Gamma_{\alpha\mu}^\beta \dot{\theta}^\mu) = 0. \quad (52)$$

Primjećujući da je

$$\dot{\theta}^\alpha \partial_\alpha \dot{\theta}^\beta = \ddot{\theta}^\beta, \quad (53)$$

uz zamjenu imena indeksa, dobivamo koordinatni oblik geodezijske jednadžbe

$$\ddot{\theta}^\alpha + \Gamma_{\mu\nu}^\alpha \dot{\theta}^\mu \dot{\theta}^\nu = 0. \quad (54)$$

U sljedećem potpoglavlju pokazujemo kako za gore diskutiranu metriku najbrže integrirati jednadžbu geodezika.

5.1 Optimalno računanje Christoffelovih simbola

Budući da rješavanje geodezijske jednadžbe značajno opterećuje računalne resurse, prvo moramo naći optimalni način računanja Christoffelovih simbola, $\Gamma_{\mu\nu}^\beta$, za metriku danu jednadžbom (40). Definijski izraz za Christoffelove simbole glasi

$$\Gamma_{\mu\nu}^\beta = \frac{1}{2} g^{\alpha\beta} (\partial_\nu g_{\mu\alpha} + \partial_\mu g_{\nu\alpha} - \partial_\alpha g_{\mu\nu}). \quad (55)$$

Ovaj generalni izraz možemo pojednostaviti koristeći izraz (40) i notaciju prethodnog poglavlja

$$\Gamma_{\mu\nu}^\beta = \frac{1}{2} g^{\alpha\beta} \left(\partial_\nu (\partial_\alpha \mathbf{f}^T \partial_\mu \mathbf{f}) \right. \quad (56)$$

$$\begin{aligned} & \left. + \partial_\mu (\partial_\alpha \mathbf{f}^T \partial_\nu \mathbf{f}) - \partial_\alpha (\partial_\mu \mathbf{f}^T \partial_\nu \mathbf{f}) \right) \\ & = \frac{1}{2} g^{\alpha\beta} \left[\partial_{\alpha\nu} \mathbf{f}^T \partial_\mu \mathbf{f} + \partial_{\mu\nu} \mathbf{f}^T \partial_\alpha \mathbf{f} \right. \\ & \quad \left. + \partial_{\alpha\mu} \mathbf{f}^T \partial_\nu \mathbf{f} + \partial_{\mu\nu} \mathbf{f}^T \partial_\alpha \mathbf{f} \right. \\ & \quad \left. - \partial_{\alpha\mu} \mathbf{f}^T \partial_\nu \mathbf{f} - \partial_{\alpha\nu} \mathbf{f}^T \partial_\mu \mathbf{f} \right], \end{aligned} \quad (57)$$

iz čega slijedi

$$\Gamma_{\mu\nu}^\beta = g^{\alpha\beta} \partial_\alpha \mathbf{f}^T \partial_{\mu\nu} \mathbf{f} \quad (58)$$

Ovaj izraz možemo jednostavnije napisati koristeći pokrate $\mathbf{J}_\mu = \partial_\mu \mathbf{f}$ i $\mathbf{H}_{\mu\nu} = \partial_{\mu\nu} \mathbf{f}$

$$\Gamma_{\mu\nu}^\beta = g^{\alpha\beta} \mathbf{J}_\alpha^T \mathbf{H}_{\mu\nu}. \quad (59)$$

Budući da ovdje izvedeni oblik Fisherove metrike sadrži Jakobijan, $\partial_\alpha f_i$, i Hessijan, $\partial_{\mu\nu} f_i$, valja primijetiti da je on jednostavniji za računalnu implementaciju od sekvencijalne primjene prvih derivacija koje bi se računale ako bismo koristili definicijski oblik Christoffelovih simbola dan izrazom (55).

6 Primjena diferencijalne geometrije na sloppy probleme

U ovome poglavlju opisujemo primjene diferencijalne geometrije u analizi sloppy problema. U prvom potpoglavlju opisujemo generalizaciju Levenberg-Marquardt metode koja uključuje gibanje po geodeziku kako bi se izbjegao rub mnogostrukosti, a u sljedećem poglavlju opisujemo MBAM metodu koja pomaže u redukciji kompliciranih modela.

6.1 Generalizacija Levenberg-Marquardt metode

Najjednostavnija primjena je generalizacija Levenberg-Marquardt metode korištenjem jednadžbe geodezika kako bi se izbjegao rub mnogostrukosti. Naime, vidimo da izraz (54) opisuje akceleraciju u prostoru parametara, $a^\alpha = \ddot{\theta}^\alpha$, koja bolje opisuje gibanje po geodeziku, dok pomak Levenberg-Marquardt metode dan jednadžbom (47) možemo shvatiti kao brzinu. Kombiniranjem ovih dviju

jednadžbi, dobivamo generalizirani pomak

$$\delta^\alpha = v^\alpha + \frac{1}{2}\Gamma_{\mu\nu}^\alpha v^\mu v^\nu, \quad (60)$$

gdje je brzina dana jednadžbom (47), t.j. $v^\alpha = g^{\alpha\mu}\mathbf{J}_\mu^T \mathbf{r}$. Ovaj izraz možemo opravdati proširujući razvoj u jednadžbi (44) do drugog reda $\mathbf{f} = \mathbf{f}_0 + \mathbf{J}_\alpha \delta^\alpha + \frac{1}{2}\mathbf{H}_{\alpha\beta} \delta^\alpha \delta^\beta$. Uz pretpostavku da je $\mathbf{H}_{\alpha\beta}$ malen, dobivamo jednadžbu

$$\mathbf{J}_\mu^T \mathbf{r} = g_{\mu\nu} \delta^\nu + \frac{1}{2}\mathbf{J}_\mu^T \mathbf{H}_{\alpha\beta} \delta^\alpha \delta^\beta. \quad (61)$$

Uzimajući u obzir da je δ malena, približno vrijedi $\delta^\mu = g^{\mu\nu}\mathbf{J}_\nu^T \mathbf{r}$, što uvrštavamo u izraz (61) gdje uz definiciju $v^\mu = \delta^\mu$ dobivamo

$$\mathbf{J}_\mu^T \mathbf{r} = g_{\mu\nu} \delta^\nu + \frac{1}{2}\mathbf{J}_\mu^T \mathbf{H}_{\alpha\beta} v^\alpha v^\beta. \quad (62)$$

Množenjem s $g^{\mu\sigma}$ dobivamo traženi izraz. U članku [2], autori su testirali efikasnost izbjegavanja ruba mnogostrukosti, gdje su pokazali da je ovaj modificirani oblik Levenberg-Marquardt metode efikasniji od tradicionalne Levenberg-Marquardt metode.

6.2 MBAM Metoda

Aproksimacijska metoda mnogostrukosti s rubom¹ (MBAM) [4] koristi se za razdvajanje onih parametara koji su doista potrebni u modelu² od onih parametara koji ne utječu bitno na konačne rezultate³. Model s manjim brojem parametara općenito će biti pogodniji za numeričke proračune, a dat će nam i bolji uvid u danu problematiku.

¹The Manifold Boundary Approximation Method.

²tzv. *stiff* parameters.

³tzv. *soft* parameters.

Model, $P(\xi, \theta)$ opisujemo kao funkciju dva seta parametara, ξ su parametri koji su bitni za opis sustava, tkzv. interesantne veličine (QoI⁴) dok su θ svi ostali parametri. Ne postoji generalna podjela parametara, ali često se za QoI uzimaju parametri za koja postoje mjerenja [5]. Postupak metode je sljedeći.

Prvi korak je nalaženje prve procjene parametara, θ_0 i pripadne Fisherove informacijske metrike, kao i njezinih svojstvenih vrijednosti.

Drugi korak uključuje integraciju geodezijske jednadžbe. Vidimo da je geodezijska jednadžba jednadžba drugog reda u τ , ali je možemo rastaviti u sustav $2N_p$ diferencijalnih jednadžbi prvog reda

$$\dot{\theta}^\alpha = y^\alpha \quad (63)$$

$$\dot{y}^\alpha = -\Gamma_{\mu\nu}^\alpha y^\mu y^\nu, \quad (64)$$

s početnim uvjetima izabranim prema točki najbolje prilagodbe i smjeru zadanom pomoću (normiranog) svojstvenog vektora, v_{min}^α koji odgovara najmanjoj svojstvenoj vrijednosti Fisherove metrike, odnosno

$$\theta^\alpha(\tau = 0) = \theta_{bf}^\alpha \quad (65)$$

$$y^\alpha(\tau = 0) = v_{min}^\alpha. \quad (66)$$

Budući da nas zanima samo kvalitativno ponašanje na rubu mnogostrukosti, visoka preciznost nije od presudne važnosti. Manja preciznost olakšava rješavanje ovog sustava diferencijalnih jednadžbi jer se vrijeme između dva koraka integracije povećava kako dolazimo bliže rubu mnogostrukosti. Po geodeziku se gibamo sve dok metrika izvrijednjena u

⁴Quantities of Interest

točkama geodezika ne postane singularna. Singularnost metrike nam ukazuje na to da je geodezik došao do ruba mnogostrukosti.

U trećem koraku iz ponašanja parametara na rubu mnogostrukosti zaključuje se kako reducirati model, što može uključivati izbacivanje kombinacije parametara koji se pokazuju nepotrebnima na rubu mnogostrukosti, te se pomoću tog uvjeta konstruira novi, jednostavniji model.

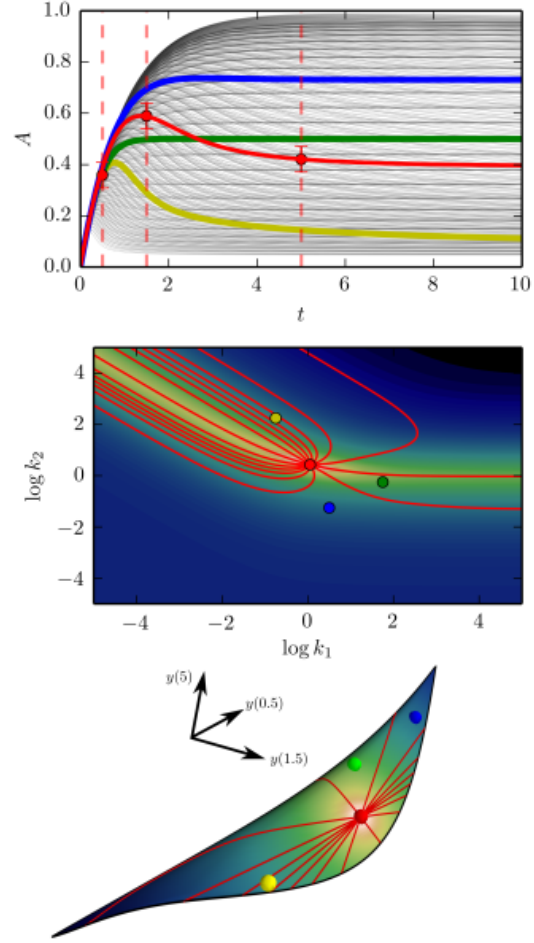
U četvrtom se koraku kalibriraju parametri novog modela, ϕ , minimizirajući informacijsku udaljenost između dva modela, tj. minimizira se KL-divergencija

$$\min_{\phi} D_{KL}(\xi, \theta || \xi, \phi) = \min_{\phi} E[\ln P(\xi, \theta) - \ln P(\xi, \phi)]. \quad (67)$$

Budući da početni model ne ovisi o ϕ , ova minimizacija se svodi na maksimizaciju log-vjerojatnosti reduciranog modela. Ovaj postupak se može ponavljati sve dok model ne postane zadovoljavajuće jednostavan. Jednostavan primjer primjene ove metode je problem biološke adaptacije koji uključuje dvije dinamičke varijable A i B i dva parametra, k_1 i k_2 opisan jednadžbama

$$\begin{aligned} \frac{dA}{dt} &= (1 - A) - k_2 AB \\ \frac{dB}{dt} &= k_1 A(1 - B), \end{aligned} \quad (68)$$

s početnim uvjetima $A(0) = B(0) = 0$. Prilagodba modela mjerenjima za različite kombinacije parametara dana je u gornjem okviru slike 4, njezin χ^2 kao funkcija parametara $k_{1,2}$ prikazan je u srednjem ok-



Slika 4: Primjer problema biološke adaptacije opisan jednadžbama (68). Prilagodba modela mjerenjima za različite kombinacije parametara dana je u gornjem okviru slike 4, gdje crvene točke označuju mjerenja. Njezin χ^2 kao funkcija parametara $k_{1,2}$ prikazan je u srednjem okviru, a na donjem okviru je prikazana mnogostrukost u prostoru triju mjerenja s pripadnim geodezicima prikazanim crvenom linijama. Slika posuđena iz [5].

viru, a na donjem okviru je prikazana mnogostrukost u prostoru triju mjerenja s pripadnim geodezicima. Sa srednjeg okvira se vidi da geodezici prate vrijednosti χ^2 funkcije, ali prelaze preko cijele mnogostrukosti, što je vidljivo s donjeg okvira.

Odaberemo li rub mnogostrukosti za koji je $k_1 \rightarrow 0$, $k_2 \rightarrow \infty$, vidimo da bi zbog divergencije jednadžbe za A bilo korisno uvesti novu varijablu $\tilde{B} = k_2 B$, te sustav postaje

$$\begin{aligned}\frac{dA}{dt} &= (1 - A) - A\tilde{B} \\ \frac{d\tilde{B}}{dt} &= k_1 k_2 A (1 - \tilde{B}/k_2).\end{aligned}\quad (69)$$

Vidimo da možemo uvesti varijablu $\tilde{k} = k_1 k_2$. U limesu $k_2 \rightarrow \infty$, $\tilde{k} = \text{const}$ dobivamo

$$\begin{aligned}\frac{dA}{dt} &= (1 - A) - A\tilde{B} \\ \frac{d\tilde{B}}{dt} &= \tilde{k}A.\end{aligned}\quad (70)$$

Iz jednadžbe (70) vidimo da smo parametar k eliminirali reskaliranjem dinamičke varijable B . Sustav opisan s \tilde{k} i \tilde{B} zadržava ponašanje na rubu mnogostrukosti kakvu je imao i početni model, ali je opisan s parametrom manje.

7 Zaključak

Opisali smo Fisherovu informacijsku metriku na prostoru parametara modela pomoću Kullback-Leibler divergencije između modela opisanih infinitezimalno bliskim parametrima kao i njezinu realizaciju na prilagodbu podataka. Nadalje, opisali smo općeniti

sloppy problem i obrazložili neke primjene Fisherove informacijske metrike u prilagodbi podataka. Opisali smo kako generalizirati standardnu Levenberg-Marquardt metodu prilagodbe podataka u slučaju mnogostrukosti s rubom. Na kraju smo opisali aproksimacijsku metodu mnogostrukosti s rubom koja se koristi u redukciji kompleksnih modela.

Literatura

- [1] J. Dobaczewski, W. Nazarewicz, and P.-G. Reinhard. Error estimates of theoretical models: a guide. *Journal of Physics G Nuclear Physics*, 41(7):074001, July 2014.
- [2] M. K. Transtrum, B. B. Machta, and J. P. Sethna. Why are Nonlinear Fits to Data so Challenging? *Physical Review Letters*, 104(6):060201, February 2010.
- [3] Charles W. Misner, Kip S. Thorne, and John Archibald Wheeler. *Gravitation*. W. H. Freeman, New York, 1973.
- [4] Mark K. Transtrum and Peng Qiu. Model reduction by manifold boundaries. *Phys. Rev. Lett.*, 113:098701, Aug 2014.
- [5] M. K. Transtrum. Manifold boundaries give “gray-box” approximations of complex models. *ArXiv e-prints*, May 2016.