Identifikace parametrů vybraných subsystémů dynamického modelu fyziologických funkcí "eGolem"

autor: Bc. Jan Šilar,
vedoucí: Mgr. Marek Mateják

Práce psána ve spolupráci s

Laboratoř biokybernetiky a počítačové podpory výuky
Ústav patologické fyziologie
1. lékařská fakulta
Universita Karlova

RNDr. Bohuslav RŮŽEK CSc - inverzní úlohy geofyziky

Identifikace modelů

- modely fyziologických dějů - simulují děje v lidském těle
- Úkol: zjistit hodnoty parametrů modelu tak, aby výsledky simulace odpovídaly datům naměřeným při experimentu na konkrétním člověku
- používáme optimalizační nebo inverzní techniky
- takto určujeme parametry jejichž měření: invazivní (zásah do těla), drahé, nemožné

Proč identifikovat parametry?

- **získání hodnot parametrů,** které je jinak obtížné zjistit $\longrightarrow$ nová diagnostická metoda využitelná v praxi
- **nafitovaný model:** model nyní simuluje konkrétního člověka $\longrightarrow$ využití v praxi (např. v anesteziologii)
- **ověření modelu:** nafitování modelu na experimentální data se zdaří $\longrightarrow$ model dobře popisuje jevy probíhající při měřeném experimentu

Model

- soustava algebraických a obyčejných diferenciálních rovnic (*DAE*)
- rovnice nelineární
- modely často hybridní - hodnoty veličin nebo celé rovnice se mění v závislosti na podmínkách (příklad s cévou)
- počet proměnných až několik tisíc

$$\Downarrow$$

S modely nejde pracovat analyticky $\Rightarrow$ numerické řešení.

- modely implementovány v simulačním jazyce *Modelica* (editor *Dymola*) a přeloženy $\rightarrow$ reziduální funkce

$$F(t, x, \dot{x}) = 0 \tag{1}$$

  jako knihovna (zatím většinou v C++)
- solver řeší model na základě reziduální funkce

Formulace úlohy

Máme:

experimentální data: $\bar{x}_E(t)$ - vektor časových průběhů veličin naměřených při experimentu na pacientovi. Každá složka – funkce času, odpovídá jedné měřené veličině.

implementovaný model: $M(\bar{p}, \bar{x}_0) \longrightarrow \bar{x}_S(t)$ - simuluje experiment, vektoru parametrů $\bar{p}$ a vektoru počátečních podmínek $\bar{x}_0$ přiřadí vektor časových průběhů simulovaných veličin $\bar{x}_S(t)$.

Úkol najít:

minimum

$$\min_{\bar{p}} \| M(\bar{p}, \bar{x}_0) - \bar{x}_E(t) \|_{L2} \qquad \text{(optimalizace)}, \qquad (2)$$

nebo řešení přeurčené soustavy rovnic pro neznámé $\bar{p}$

$$M(\bar{p}, \bar{x}_0, t_i) = \bar{x}_E(t_i), \ \forall i \in N \qquad \text{(inverze)}. \qquad (3)$$

Lokální / globální metody

Účelová funkce může mít na prohledávaném intervalu jedno nebo více lokálních minim.

- **jedno lokální minimum** $\Rightarrow$ použijeme lokální metodu (např. simplexovou) - rychlý výpočet.
- **více lokálních minim** $\Rightarrow$ globální metody. Pro n parametrů prohledáváme prostor dimense n. Hrubá síla $O(exp(n))$.
- fyziologické modely - většinou více lokálních minim $\rightarrow$ zabýváme se hlavně globálními metodami
- vyhodnocování modelu časově náročné $\rightarrow$ efektivní jsou metody, které minimalizují počet vyhodnocení modelu

Odhad počtu lokálních minim

kvůli výběru metody

- náhodně generujeme počáteční odhady $\bar{p}$
- voláme opakovaně lokální minimalizační metodu (simplexovou)
- N numerických aproximací minim $\longrightarrow$ shluková analýza
- shlukujeme do $m =$1, 2, 3 ... shluků
- sledujeme hodnotu kriteriální funkce

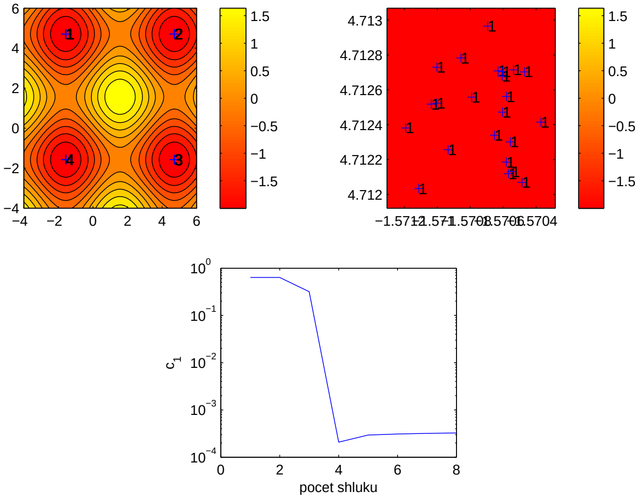
$$c = \sum_{i=1}^{m} \sum_{j=1}^{n} \mathrm{std}(p_j^i), \tag{4}$$

n je počet parametrů, m je počet shluků. Součet směrodatných odchylek přes všechny parametry a všechny shluky.

Příklad:

$$y = \sin(x) + \sin(y) \tag{5}$$

na intervalu $< -4, 6 > \times < -4, 6 >$, kde má 4 minima.



Funkce přestane prudce klesat → správný počet shluků

Testované metody

- ANNI
- Isometrická metoda (obě pro inverzní metody navržené pro úlohy geofyziky)
- CMA-ES - evoluční algoritmus

Funkce Matlabu (balíčky *Optimization Toolbox* a *Global Optimization Toolbox*)

- `fminsearch` - simplexová metoda
- `ga` - genetický algoritmus
- `patternsearch`
- `simulannealbnd` - simulované žíhání

Testy

fitování dvou testovacích funkcí, jedna z nich:

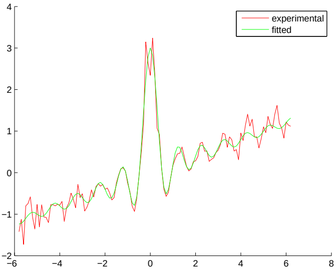
$$y = c \cdot \mathrm{sinc}(a \cdot (x - b)) + d \cdot x \tag{6}$$

$a = 2,\ b = 0\ ,\ c = 3,\ d = 0.2$

- Vygenerujeme hodnoty funkce na nějaké síťce
- Data zašumíme
- "Zapomeneme" hodnoty parametrů a fitujeme

Metoda	N_{eval}	err_{sum}
ANNI	809	$5.6 \cdot 10^{-17}$
CMA-ES	–	–
GA	1041	0.047
Isometric	6796	0.043
PatternSearch	683	0.043
SimmulAnneal	2559	0.039

ANNI – nejpřesnější
PatternSearch – nejméně vyhodnocení
CMA-ES – nefungovala.



Model odporu krevního řečiště při zátěži - úvod

Experiment

- data od týmu Dr. Uwe Hoffmanna
- pacient na rotopedu, zátěž 30W, 200W, 0W
- měřeno *HR* – srdeční frekvence, *MAP* – střední arteriální tlak, *CO* – minutový srdeční tok, *PW* – celkový výkon.

Fyziologie

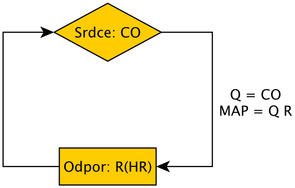
- při fyzické zátěži – svaly spotřebovávají kyslík
- dostatek kyslíku ve svalech → změny odporu částí krevního řečiště – redistribuce krve (kyslíku) do jednotlivých tkání
- způsobeno *sympatickou reakcí* – dva efekty
  - stažení cév vegetativních orgánů → méně krve do orgánů, více krve do svalů, zvětšení celkového odporu.
  - zvýšení *HR*
- *HR* měřená – ukazatel velikosti sympatické reakce

Základní model

Zjednodušené krevní řečiště modelu eGolem. Nepočítá každý úder srdce, jen střední hodnoty toků.

Analogie s elektrickým obvodem.

- srdce – pumpa, dodává tok CO (měřená veličina).
- krevní řečiště má jediný odpor – celkový hydraulický odpor velkého krevního oběhu
  $R = R_0 \cdot q(HR, \tau)$
- R_0 základní systémový odpor (bez simpatické reakce),
- q koeficient změny odporu vlivem simpatické reakce, závisí na HR, τ - zpoždění změny (stažení nenastává okamžitě)



Identifikace základního modelu

- Identifikujeme parametry $R_0 \in \langle 0, 300 \rangle$ a $\tau \in \langle 0, 20 \rangle$.
- Účelová funkce má na intervalu jediné lokální minimum.

Metoda	N_{eval}	(R_0, τ)	f_{val}
ANNI	1613	$(172.6, 9.8)$	$4.6914 \cdot 10^3$
CMA-ES	2379	$(187.0, 9.8)$	$4.6873 \cdot 10^3$
GA	2021	$(11.6, 9.4)$	$1.1383 \cdot 10^4$
Isometric	3117	$(182.3, 9.8)$	$4.6881 \cdot 10^3$
PatSe	414	$(187.0, 9.8)$	$4.6873 \cdot 10^3$
SimAn	6002	$(166.1, 9.8)$	$4.6898 \cdot 10^3$
Simplex	83	$(188.0, 9.8)$	$4.6873 \cdot 10^3$

Simplexová – nejefektivnější,
PatternSearch – druhá,
GA – nenašla.

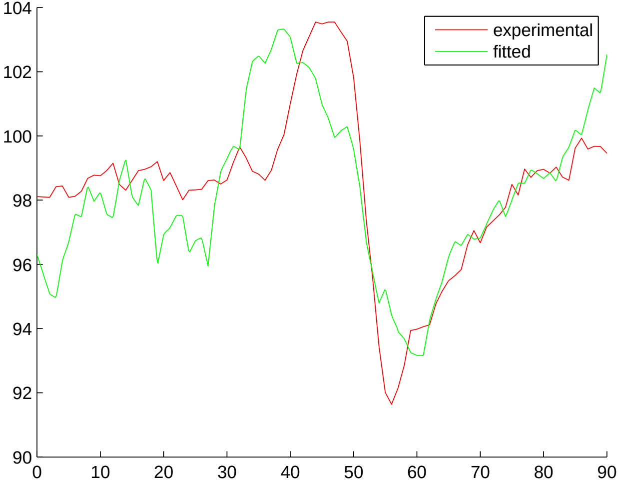
Chyba vlivem rozdílu křivek, těžko porovnat přesnost.

Rozšířený model

- Další vliv – lokální svalová reakce.
- Cévy přivádějící krev do pracujících svalů se samy rozšíří $\longrightarrow$ více krve do svalů, snížení celkového odporu.
  $R = R_0 \cdot \frac{q(HR,\tau)}{rf(PW,\tau_{inc},c_{dec})}$
- rf relaxační faktor – závisí na výkonu svalů, τ_{inc} zpoždění náběhu relaxace, c_{dec} koeficient ústupu reakce

Metoda	N_{eval}	$(R_0, \tau, \tau_{inc}, c_{dec})$	f_{val}
ANNI	1617	$(66.8, 9.84, 4.24, 5.51)$	269.1
CMA-ES	1995	$(190.0, 10.0, 1.72 \cdot 10^5, 2.91 \cdot 10^5)$	$4.69 \cdot 10^3$
GA	1701	$(51.2731, 9.86, 4.365.36)$	275.0
Isometric	3434	$(69.6, 9.89, 4.03, 5.29)$	252.4
PatternSearch	1863	$(71.2, 9.89, 4.13, 5.34)$	251.8
SimmulAnneal	12002	$(182.7, 9.97, 3.45, 5.31)$	318.6095
Simplex	387	$(55.49.8818.311.8)$	$2.17 \cdot 10^3$

ANNI – nejefektivnější
PatternSearch – nejpřesnější
Simulované žíhání – neefektivní, nepřesné
CMA-ES, Simplexová – špatné výsledky



Experimentální a nafitovaná data rozšířeného modelu si již docela dobře odpovídají. Posunutá maxima – možnost upřesnit model.

Projekty, spolupráce

Současné:

- CESNET - vybudování gridu pro identifikaci parametrů (výpočetně náročné). Vytvoření robustního rozhraní mezi modelem a identifikačním algoritmem pro paralelizaci.
- Spolupráce s týmem Dr. Uwe Hoffmannan, Sportovní univerzita v Kolíně nad Rýnem. Fyziologie v extrémních podmínkách - vrcholoví sportovci, potapěči, kosmonauti. Zpracováváme jejich data.

Plánované:

- GAČR - Podán grantový návrh. Identifikace parametrů modelu krevního řečiště. Plánováno několik experimentů – měření mnoha veličin.

Co se udělalo

- vybrány a otestovány vhodné metody pro identifikaci fyziologických modelů.
- vytvořeno prostředí pro identifikaci modelů. Umožňuje rychlé přepínání identifikovaných modelů, snadná volba identifikační metody a parametrů metody. Zpřehledňuje práci. Implementace rozhraní mezi identifikačním algoritmem a modelem pomocí tcp-ip.
- navržena metoda pro zjištění počtu minim (hlavně 1 nebo více)
- pospána citlivostní analýza MC simulace chyb, implementace po odevzdání.
- identifikační prostředí použito při identifikaci modelu odporu krevního řečiště při zátěži → navrženo rozšíření modelu eGolem.

Co se plánuje udělat

- zpřehlednit kód.
- robustnější rozhraní pro komunikaci identifikační algoritmus – model.
- paralelizace vybraných metod – grid, později CUDA.
- hybridní přístup $\longrightarrow$ lokální metody.
- identifikační prostředí součástí editoru modelů.

Děkuji za pozornost.

Použití aproximačních modelů v medicíně

- **kdy se používají:**
  mnoho vztahů v používaných v modelech je je získáno aproximací měřených dat nějakou funkcí. Fyziologické a fyzikální mechanizmi jsou často složité nebo neznámé.
- **při identifikaci parametrů**
  - nemáme k dispozici "trénovací množinu" – množinu dvojic $(\bar{p}_i, \bar{d}_i)$ pro naučení/nastavení aproximačního modelu.
  - jediná možnost nastavit aproximační model tak, aby kopíroval fyziologický model $\rightarrow$ mnoho vyhodnocení dopředného modelu.
  - Aproximační model – rychlejší výpočet než fyziologický $\rightarrow$ vyplatí se, pokud bychom prováděli identifikaci opakovaně, s různými daty.
- Neuronové sítě – zpracování a rozpoznávání signálů. Automatická diagnóza onemocnění, detekce nádoru v obraze (MR), zpracování EKG a EEG...

Citlivostní analýza a závislost parametrů

Citlivost identifikovaných parametrů na chyby v datech.
Závíslost parametrů – různé kombinace hodnot parametrů vedou k podobnému výsledeku modelu $\longrightarrow$ parametry není možné identifikovat.

MC simulace chyb

- identifikované parmetry $\longrightarrow$ modelová data.
- opakujeme:
  - generování náhodného šumu (zde gaussovský $\sigma^2 = 1$, $\mu = 0$)
  - data + šum
  - identifikace $\longrightarrow$ parametry
- spočtemem kovariančí matici realizací parametrů

Výsledky simulace chyb

$$\bar{p} = (R_0, \tau, \tau_{inc}, c_{dec}) = (71.2, 9.89, 4.13, 5.34) \tag{7}$$

$$cov(pp) = \begin{pmatrix} 5.8309 & 0.0088 & -0.1496 & -0.0330 \\ 0.0088 & 0.0000 & -0.0004 & -0.0001 \\ -0.1496 & -0.0004 & 0.0159 & 0.0034 \\ -0.0330 & -0.0001 & 0.0034 & 0.0015 \end{pmatrix} \tag{8}$$

Největší rozptyl má první parametr, ale

$$err \approx \sigma = \sqrt{\sigma^2} = (2.41473,\ 0.0,\ 0.1261,\ 0.03873) \tag{9}$$

$$err_r \approx \sigma/\bar{p} = (0.034\ , 0.0,\ 0.031,\ 0.0090) \tag{10}$$

Kdyby byl rozptyl chyby měření dat roven 1, odpovídaly by chyby identifikovaných parametrů zhruba těmto hodnotám.
Mimodigonální prvky jsou relativné malé, parametry by neměly být závislé.

Děkuji za pozornost.