# *PA054: Formální modely v systémové biologii*

David Šafránek

17.5.2013

# Obsah

# Obsah

# Estimace parametrů

- co je cílem?
  - najít takovou valuaci parametrů, která nejlépe odpovídá experimentálně zjištěným time-course datům
  - chceme tedy co nejvíce přiblížit simulaci experimentálním datům (tzv. **fitting**)
  - tradiční pojetí pro spojitou sémantiku
- lineární regrese požaduje normální rozložení chyb měření
- transformací se nepřesnosti kumulují
- model je inherentně nelineární
  - nelineární jsou i naměřená data
- problém fittingu chápán jako optimalizační problém
- mnoho heuristických metod pro aproximativní řešení

# *Problém reverzního inženýrství*

- tzv. **inverzní problémy**
  - cílem je získat model z pozorování systému
  - obecně řešeno v teorii systémů (identifikace systémů)
  - pro nelineární systémy obecně neřešitelné
  - viz. IV120
- obecné schema řešení inverzního problému:
  1. identifikace vztahů mezi proměnnými
  2. identifikace funkcí popisujících sémantiku jednotlivých vztahů (např. zákon záchování hmoty, Michaelis-Menten, Hill, ...)
  3. **estimace hodnot parametrů ve funkcích získaných v předch. bodě**

# *Estimace parametrů optimalizací*

- obecný postup:
  1. srovnej experimentální time-course se simulovaným time-course
  2. pokud rozdíl menší než nastavená tolerance $\rightarrow$ DONE
     jinak modifikuj parametry modelu
  3. proveď time-course simulaci modelu
  4. iteruj (1)

# Estimace parametrů optimalizací

- mějme model daný systémem $\frac{dx}{dt} = f(x, p)$ kde $x$ je stavový vektor a $p$ je **vektor hodnot parametrů**
- uvažujme $T = \langle t_1, ..., t_m \rangle$ rostoucí posloupnost časových bodů (tzv. **časovou řadu**)
- předpokládejme posloupnost $\langle x(t_1), ..., x(t_m) \rangle$ je aproximace řešení $x(t)$ zachycená v časové řadě $T$ (**simulace**)
- zdůrazněme fakt, že simulace byla získána při nastavení hodnot parametrů $p$, označením $x(t)_p$
- mějme **experiment** jako posloupnost vektorů naměřených veličin $\langle y(t_1), ..., y(t_m) \rangle$ v časové řadě $T$
- pro jednoduchost uvažujme $dim(x) = dim(y) = 1$ (obecně $dim(x) \geq dim(y)$ libovolné, ale složitější formulace)

# *Estimace parametrů optimalizací*

- definujeme odchylku experimentu od simulace v časovém bodě $t_i$ jako tzv. **reziduál**:

$$r(t_i, p) = y(t_i) - x(t_i)_p$$

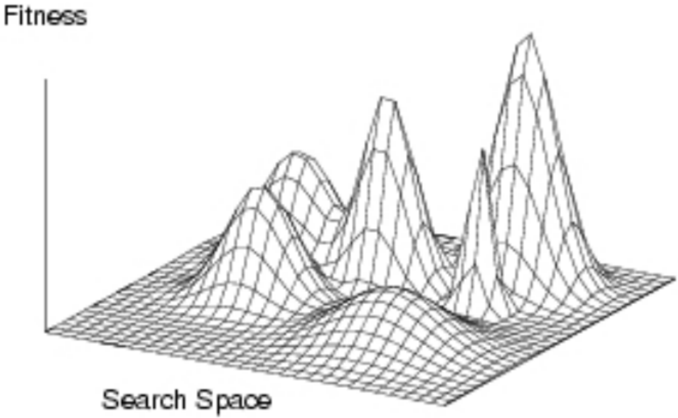
- reziduál chápeme jako funkci závislou na nastavení parametrů simulovaného modelu
- srovnání experimentu a simulace je vyjádřeno jako součet čtverců reziduálů přes vš. časové body $T$:

$$\mathcal{S}(p) = \sum_{i}^{m} (r(t_i, p))^2$$

# *Estimace parametrů optimalizací*

- funkce $\mathcal{S}(p)$ se nazývá **užitková funkce**
- vystihuje průměrnou odchylku simulace od experimentu přes danou časovou řadu
- **minimální hodnota** $\mathcal{S}(p)$ určuje optimální vektor hodnot parametrů $p$, který globálně minimalizuje rozdíl mezi experimentem a modelem
- jedná se o nelineární funkci
- počet neurčitých parametrů určuje její dimenzi

# *Optimalizační krajina*



# *Procházky po optimalizační krajině...*

- CÍL: najít **globální minimum**
- nejpoužívanější jsou stochastické black-box přístupy:
    - náhodné procházení (random search)
    - evoluční strategie (evolution strategy)
    - ...
- black-box znamená absolutní nezávislost na tvaru užitkové funkce
- existují i metody, které využívají znalosti užitkové funkce (např. simulované žíhání, Truncated Newton, ...)

# *Procházky po optimalizační krajině...*

## *Random Search*

1. inicializuj náhodně výchozí hodnotu $p$
   - typicky z rovnoměrného rozložení
2. dokud není překročen povolený počet iterací, prováděj:
   2.1 sampluj novou pozici $p'$
   $\rightarrow$ uniformní náhodný výběr z hyperkoule o daném poloměru
   2.2 spočítej $\mathcal{S}(p')$
   2.3 pokud $\mathcal{S}(p') < \mathcal{S}(p)$, nastav novou pozici $p := p'$
3. $p$ nastaveno na nejvýhodnější pozici (z pohledu běhu algoritmu)

*Pozn. Existují varianty s fixním i adaptivním poloměrem.*

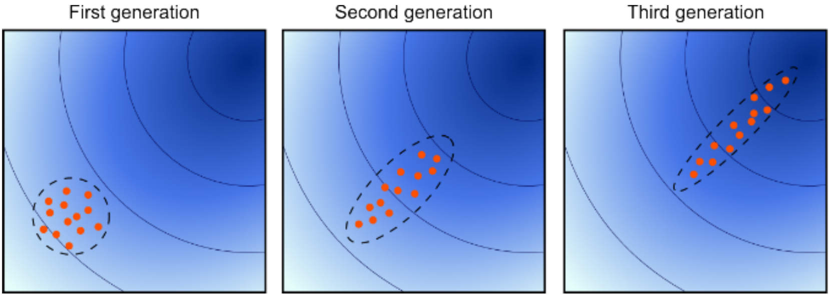
# *Procházky po optimalizační krajině...*

## *Evoluční strategie*

- inherentně adaptivní metoda, redukuje počet
- staví na výběrech z (vícerozměrného) normálního rozložení
- rozměr daný dimenzí vektoru parametrů
- značeno CMA-ES (Covariance Matrix Evolution Strategy)
- postup:
  1. vytvoř generaci
     $\rightarrow$ sampluj rozložení pozic hodnot $P$ dle normálního rozložení
  2. pro každé $p \in P$ spočítej $\mathcal{S}(p)$
  3. adaptace: uprav parametry normálního rozložení pro další iteraci
     $\rightarrow$ různé varianty, mohou být velice komplexní adaptace citlivé na charakter evoluční krajiny – adaptace kovarianční matice
- metoda se ukazuje výhodná pro biologické modely (vysoká míra neznalosti parametrů)

# *Procházky po optimalizační krajině...*

## *Evoluční strategie*



# Tuning dynamických modelů

## Estimace parametrů vzhledem k hypotézám

- estimaci parametrů lze rozšířit pro temporální specifikace (hypotézy)
  - každý experiment lze zakódovat jako LTL formuli
  - typické abstraktnější vlastnosti dávají větší stupeň volnost
- mějme specifikaci chování reprezentovanou LTL formulí $\varphi$
- uvažme parametrizovaný systém $\frac{dx}{dt} = f(x, p)$
- uvažme časovou řadu $T$
- $x_{p,T}$ značí simulovanou posloupnost v časové řadě $T$ s nastavením parametrů $p$
- definujme užitkovou funkci

$$\mathcal{S}(p) = vd(x_{p,T}, \varphi)$$

, kde $vd(x_{p,T}, \varphi)$ určuje míru splněnosti formule $\varphi$ na $x_{p,T}$

# Aplikace formálních metod

- temporální vlastnosti jako dynamická omezení
    - obecnější než časové řady
    - časové řady je také možné vyjádřit
    - logiky interpretované na spojitých trajektoriích (QLTL, STL)
- optimalizace vůči míře splněnosti
- kritériem je maximalizovat robustnost splnitelnosti vlastnosti
- nástroje Breach (VERIMAG), BioCHAM (INRIA), Parasim (FIMU)

# *Kvalitativní modely*

- parametrizována je přechodová relace
- např. boolovské modely regulačních sítí
- lze použít symbolické metody (např. NuSMV)
- explicitní metody (SimBioNet)
- heuristiky pro explicitní metody (Parsybone)

# *Stochastické modely*

- spojitý prostor parametrů (rates)
- problém parametrizace CTMC
- metody tzv. momentů (numerická simulace charakteristických funkcí pravd. rozložení)
- numerické metody vycházející z tranzientní analýzy (výpočetně náročné, ve vývoji)