HelmholtzZentrum münchen
German Research Center for Environmental Health

Evaluation of Derivative-Free Optimizers for Parameter Estimation in Systems Biology

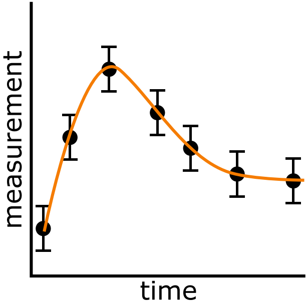
Yannik Schälte, Paul Stapor, Jan Hasenauer

FOSBE Chicago, 2018-08-07

Helmholtz Zentrum München, Institute of Computational Biology
Technische Universität München, Department of Mathematics

Motivation

Parameter estimation

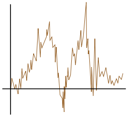


For an objective $J : \mathbb{R}^n \to \mathbb{R}$ find

$$\min_\theta J(\theta)$$

subject to $\ell \leq \theta \leq u$.

Derivative-free optimization: Why?



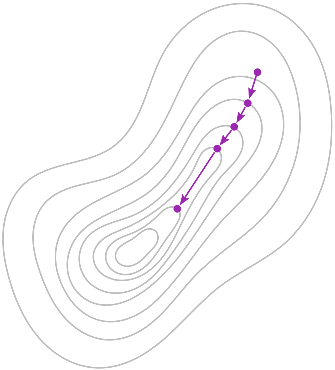
- J is given as a black-box
- J is noisy, costly, or non-smooth
- J is multi-modal
- simple to use

Optimizers

Local methods

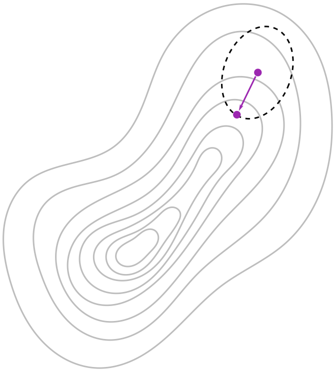
- FMINCON
- IMFIL
- BOBYQA
- PATTERNSEARCH-GPS
- PATTERNSEARCH-MADS
- RCS
- DHC
- MEIGO-DHC
- FMINSEARCHBND
- SIMULANNEALBND

Local methods



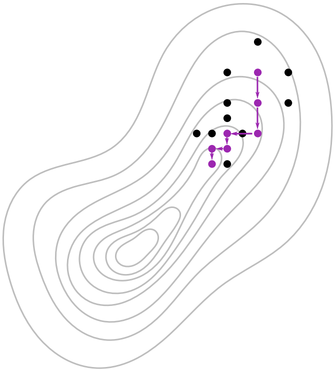
- **FMINCON**
- IMFIL
- BOBYQA
- PATTERNSEARCH-GPS
- PATTERNSEARCH-MADS
- RCS
- DHC
- MEIGO-DHC
- FMINSEARCHBND
- SIMULANNEALBND

Local methods



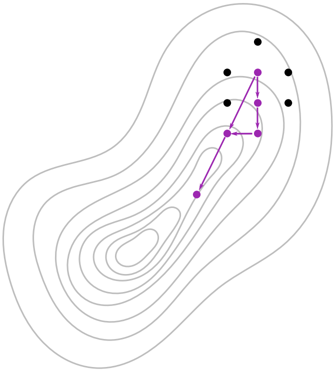
- FMINCON
- IMFIL
- **BOBYQA**
- PATTERNSEARCH-GPS
- PATTERNSEARCH-MADS
- RCS
- DHC
- MEIGO-DHC
- FMINSEARCHBND
- SIMULANNEALBND

Local methods



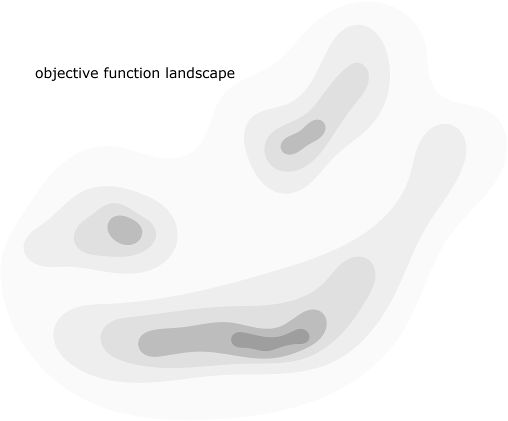
- FMINCON
- IMFIL
- BOBYQA
- PATTERNSEARCH-GPS
- PATTERNSEARCH-MADS
- RCS
- DHC
- MEIGO-DHC
- FMINSEARCHBND
- SIMULANNEALBND

Local methods

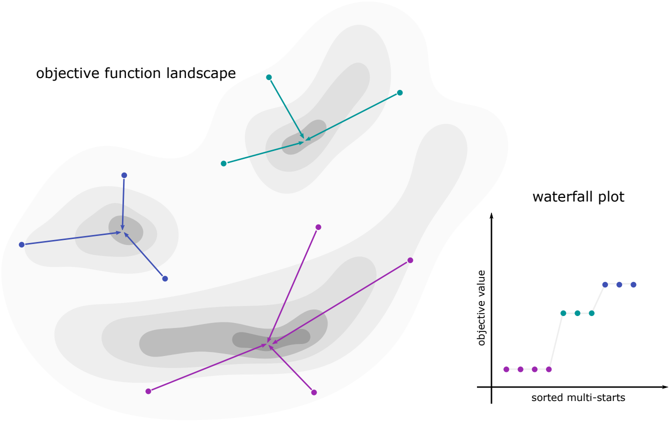


- FMINCON
- IMFIL
- BOBYQA
- PATTERNSEARCH-GPS
- PATTERNSEARCH-MADS
- RCS
- **DHC**
- MEIGO-DHC
- FMINSEARCHBND
- SIMULANNEALBND

Multi-start local optimization



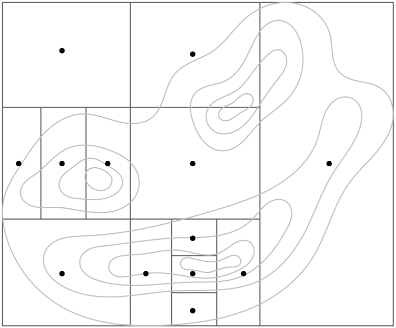
Multi-start local optimization



Global methods

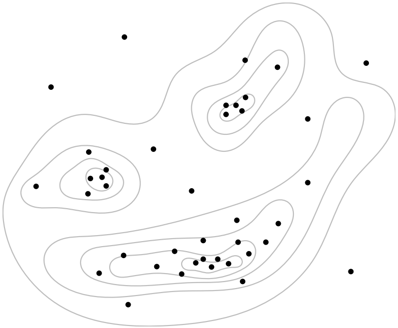
- DIRECT
- MCS
- CMAES
- GA
- PARTICLESWARM
- PSWARM
- MEIGO-ESS-BOBYQA
- MEIGO-ESS-DHC
- MEIGO-ESS-MEIGO-DHC

Global methods



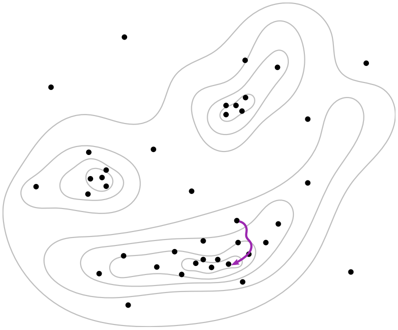
- **DIRECT**
- **MCS**
- CMAES
- GA
- PARTICLESWARM
- PSWARM
- MEIGO-ESS-BOBYQA
- MEIGO-ESS-DHC
- MEIGO-ESS-MEIGO-DHC

Global methods



- DIRECT
- MCS
- **CMAES**
- **GA**
- **PARTICLESWARM**
- **PSWARM**
- MEIGO-ESS-BOBYQA
- MEIGO-ESS-DHC
- MEIGO-ESS-MEIGO-DHC

Global methods

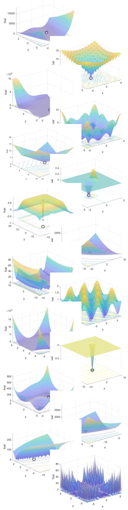


- DIRECT
- MCS
- CMAES
- GA
- PARTICLESWARM
- PSWARM
- **MEIGO-ESS-BOBYQA**
- **MEIGO-ESS-DHC**
- **MEIGO-ESS-MEIGO-DHC**

Evaluation on classic test problems

- cover issues optimizers typically have problems with
- parameters: 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 200, 300

	smooth	non-smooth	total
convex	130	65	195
non-convex	157	114	271
total	287	179	466



Experimental setup on classic test problems

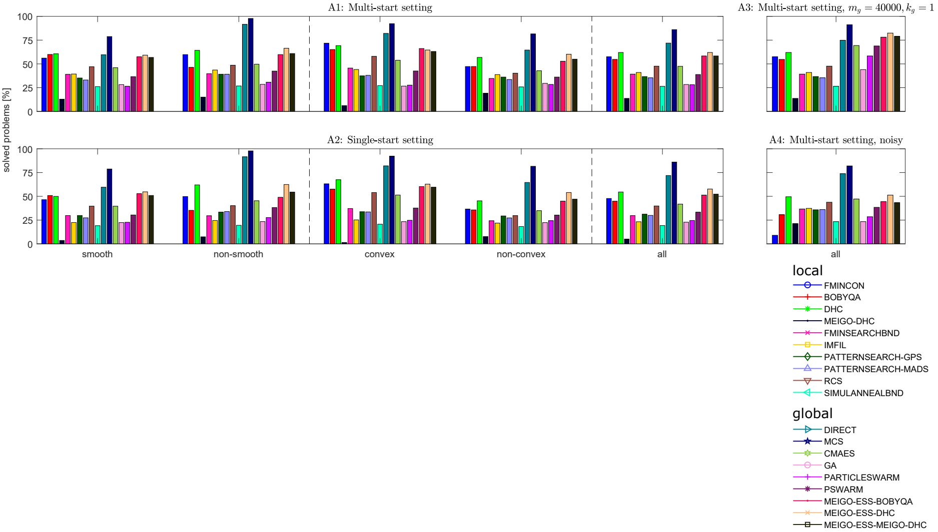
A1 Run all optimizers $k = 20$ with budgets of $m = 2000$ function calls. A problem is regarded solved if the best function value J found by the k satisfies $J \leq J^* + \varepsilon$.

A2 Like A1, but consider every of the k runs on its own to assess reliability.

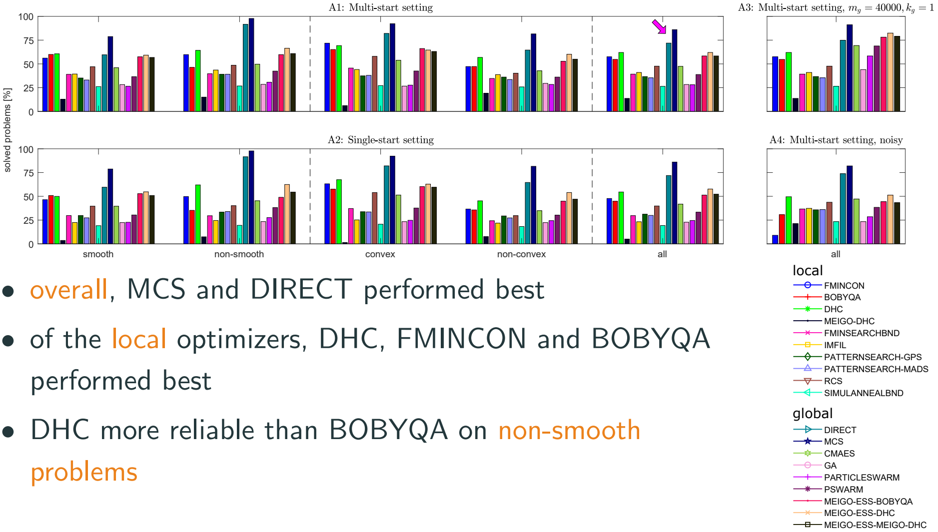
A3 Like A1, but run global optimizers $k = 1$ time with a budget of $m = 40000$ function calls.

A4 Like A1, but Gaussian noise added to every function call.

Results for classic test problems

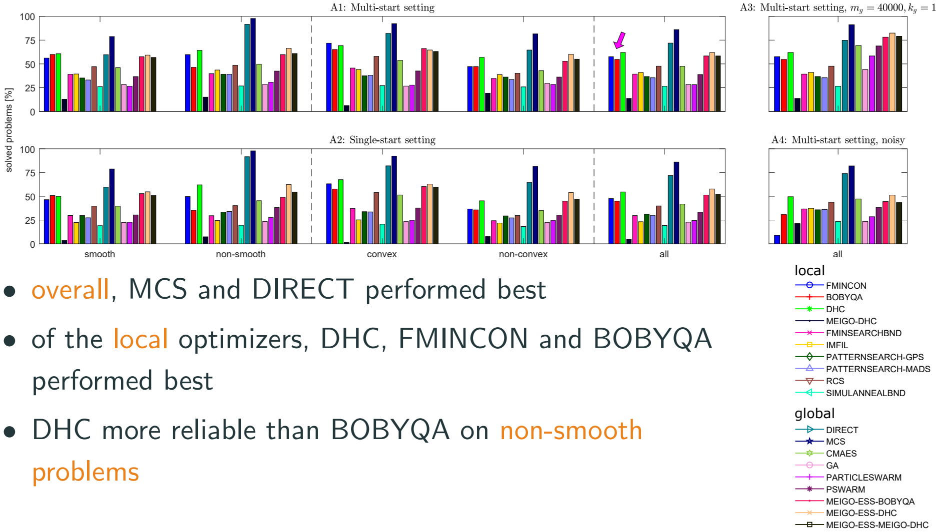


Results for classic test problems



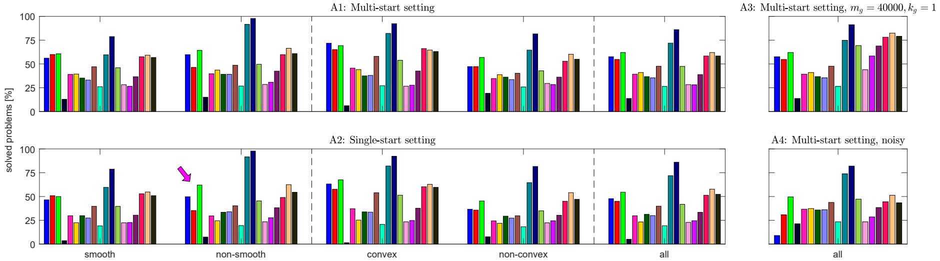
- overall, MCS and DIRECT performed best
- of the local optimizers, DHC, FMINCON and BOBYQA performed best
- DHC more reliable than BOBYQA on non-smooth problems

Results for classic test problems

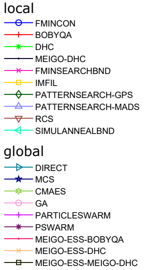


- overall, MCS and DIRECT performed best
- of the local optimizers, DHC, FMINCON and BOBYQA performed best
- DHC more reliable than BOBYQA on non-smooth problems

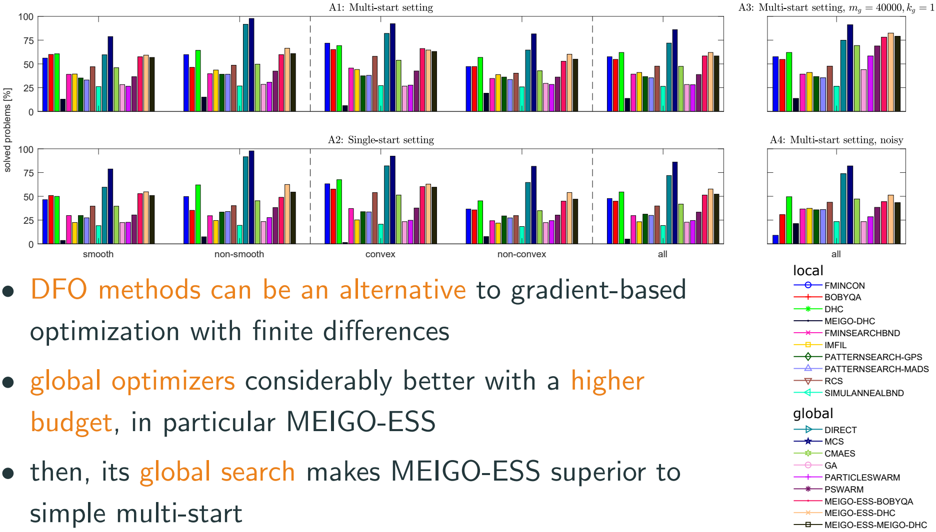
Results for classic test problems



- overall, MCS and DIRECT performed best
- of the local optimizers, DHC, FMINCON and BOBYQA performed best
- DHC more reliable than BOBYQA on non-smooth problems

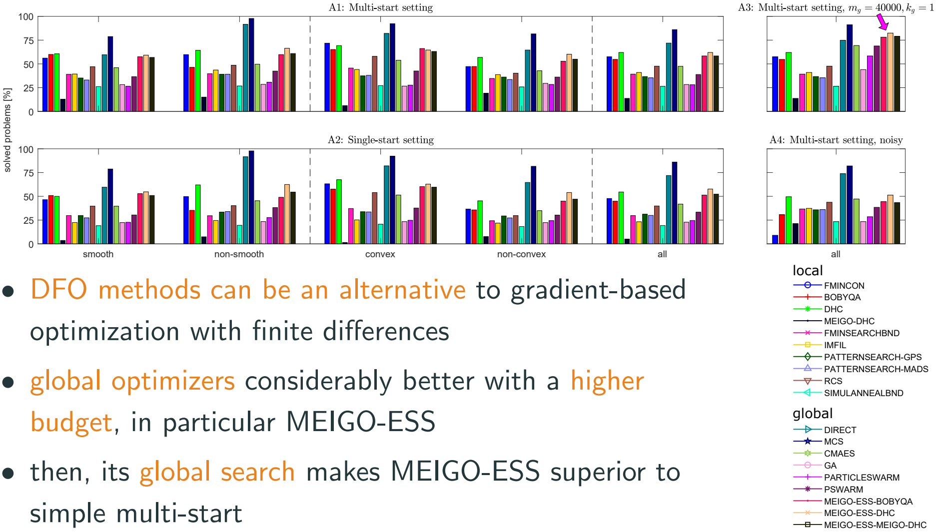


Results for classic test problems



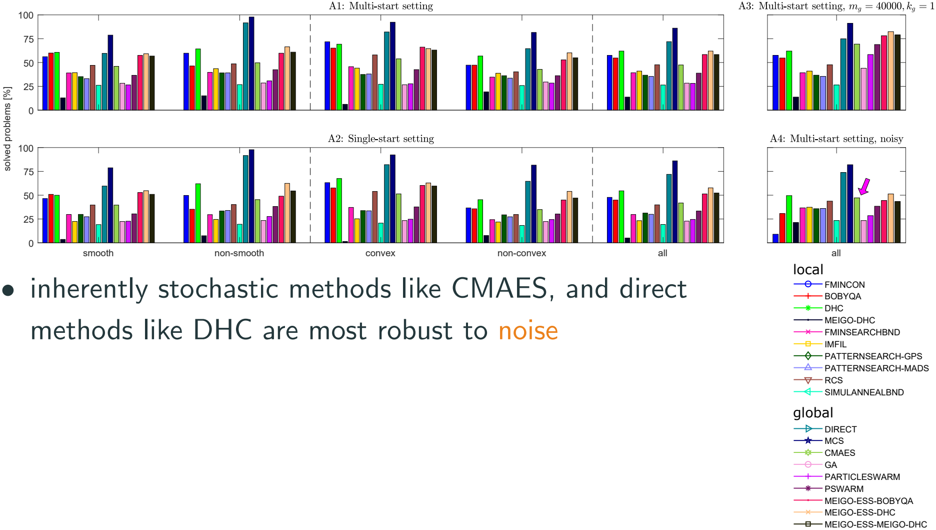
- DFO methods can be an alternative to gradient-based optimization with finite differences
- global optimizers considerably better with a higher budget, in particular MEIGO-ESS
- then, its global search makes MEIGO-ESS superior to simple multi-start

Results for classic test problems



- DFO methods can be an alternative to gradient-based optimization with finite differences
- global optimizers considerably better with a higher budget, in particular MEIGO-ESS
- then, its global search makes MEIGO-ESS superior to simple multi-start

Results for classic test problems



- inherently stochastic methods like CMAES, and direct methods like DHC are most robust to noise

Performance on biological ODE models

Biological ODE models

ID	Description	Parameters	Source
M1	conversion reaction	2	-
M2	enzyme-catalyzed reaction	4	-
M3	mRNA transfection	5	Leonhardt et al. 2014
M4	Pom1 gradient formation	7	Hross et al. 2016
M5	Hopf bifurcation	11	Ballnus et al. 2017
M6	JAK-STAT signaling	17	Swameye et al. 2003
M7	RAF-MEK-ERK signaling	28	Fiedler et al. 2016
M8	histone methylation	48	Zheng et al. 2012

- J: negative log-likelihood assuming Gaussian measurement noise

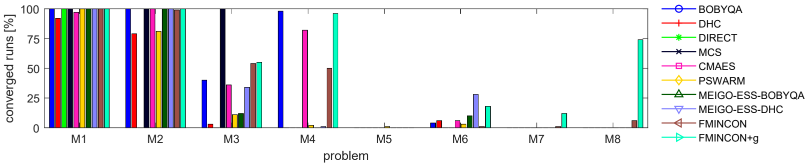
Experimental setup on biological ODE models

- best optimizers from first study
- FMINCON+g for comparison, i.e. with analytic gradients
- $k = 100$ starts, function call budgets based on prior experience
- used the toolboxes AMICI[1] and PESTO[2]

[1] Fröhlich et al.. Scalable parameter estimation for genome-scale biochemical reaction networks. Plos Computational Biology 2017

[2] Stapor et al.. PESTO: Parameter EStimation TOolbox. Bioinformatics 2018

Performance on biological ODE models

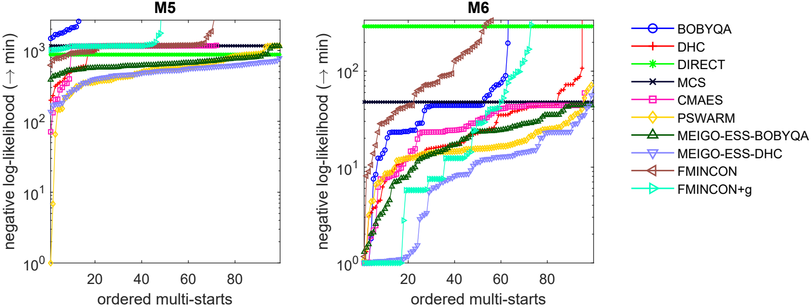


	solved problems
BOBYQA	5
DHC	4
DIRECT	1
MCS	3
CMAES	5
PSWARM	6
MEIGO-ESS-BOBYQA	4
MEIGO-ESS-DHC	4
FMINCON	7
FMINGON+g	7

- FMINCON, FMINCON+g performed best
- marked contrast to classic test functions: DFO methods only reliable on simple models
- analytic gradients improved convergence considerably

Performance on biological ODE models

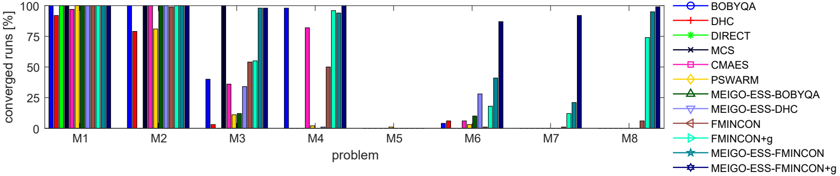
Waterfall plots



Hybrid optimizers

... analytic gradients improve convergence

- idea: combine global DFO methods with local derivative-based searches



- study: profound comparison of MEIGO-ESS and multi-start[1]

[1] Villaverde et al.. Benchmarking optim. methods for parameter estimation in large kinetic models. 2018 (submitted)

Conclusion

Conclusion

- know your problem
- derivative-based optimization superior to DFO for ODE models
- if possible, compute derivatives
- exceptions confirm the rule

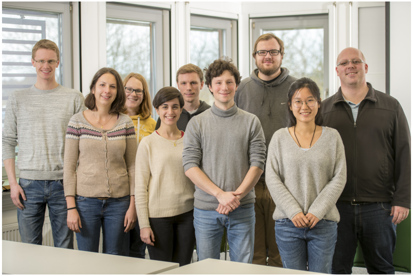
Not everything is a nail.

`github.com/icb-dcm`

Schälte, Stapor, Hasenauer. Evaluation of Derivative-Free Optimizers for Parameter Estimation in Systems Biology. FOSBE 2018

Acknowledgments

- ICB-DCM:
  - Paul Stapor
  - Fabian Fröhlich
  - Daniel Weindl
  - Jan Hasenauer
- IIM-CSIC:
  - Julio Banga
  - Jose Egea
- the toolbox authors
- ... and many more



Maximum likelihood estimation

$$\bar{y}(t) = y(t, \theta) + \varepsilon(\theta)$$

$$J(\theta) = -\log(p(\theta|\bar{y})) - \log(p(\theta))$$

Efficient computation of derivatives for ODE models

- $\nabla_\theta J$ can be computed for ODE models using forward (FSA) or adjoint sensitivity analysis (ASA)
- implemented in the toolbox AMICI
- in particular valuable for large-scale models

Forward sensitivity analysis for ODE models

State and output sensitivities

$$s^x(t,\theta) := \nabla_\theta x(t,\theta) \quad s^y(t,\theta) := \nabla_\theta y(t,\theta)$$

Forward sensitivity equation

$$\begin{aligned} \dot{s}^x &= \frac{\partial f}{\partial x} s^x + \frac{\partial f}{\partial \theta}, \qquad s^x(0,\theta) = \frac{\partial x_0}{\partial \theta}(\theta) \\ \dot{s}^y &= \frac{\partial g}{\partial x} s^x + \frac{\partial g}{\partial \theta} \end{aligned}$$

Augment and solve ODE of dim $n_x(1+n_\theta)$

Adjoint sensitivity analysis for ODE models

1. Calculation of **state** via forward simulation

$$\begin{aligned} \dot{x}(t,\theta) &= f(x(t,\theta),\theta) \\ \dot{y}(t,\theta) &= g(x(t,\theta),\theta) \end{aligned}$$

2. Calculation of **adjoint state** $p(t) \in \mathbb{R}^{n_x}$ as solution to backward ODE

$$\lim_{t \to t_N^+} p(t) = 0$$

$$\text{for } j = n_t : -1 : 1$$

$$\dot{p}(t) = -\frac{\partial f}{\partial x}\Big|_{x(t,\theta),\theta}^T p(t), \quad t \in (t_{j-t}, t_j)$$

$$\text{with } p(t_j) = \lim_{t \to t_j^+} + \sum_{k=1}^{n_y} \frac{\partial g_k}{\partial x}\Big|_{x(t_j,\theta),\theta} \frac{\bar{y}_{jk} - y_k(t_j,\theta)}{\sigma_{jk}^2}$$

3. Calculation of **gradient** via 1-dim integral

$$\nabla_\theta J(\theta) = -\int_0^{T_N} p(t)^T \frac{\partial f}{\partial \theta}\Big|_{x(t,\theta),\theta} dt - p(t_0)^T \frac{\partial x_0}{\partial \theta}\Big|_\theta - \sum_{j,k} \frac{\partial g_k}{\partial \theta}\Big|_{x(t_j,\theta),\theta}^T \frac{\bar{y}_{jk} - y_k(t_j,\theta)}{\sigma_{jk}^2}$$