HelmholtzZentrum münchen
German Research Center for Environmental Health

RHEINISCHE FRIEDRICH-WILHELMS-UNIVERSITÄT BONN

Workshop on Computational Models in Medicine and Biology, Braunschweig

Tutorial on

Mechanistic modelling of (large-scale) biochemical reaction networks

Jan Hasenauer

Faculty of Mathematics and Natural Sciences, University of Bonn

Research Group for Data-driven Computational Modelling, Helmholtz Center Munich

HELMHOLTZ ASSOCIATION

There are many different types of models

Model ‘models’

Car ‘models’

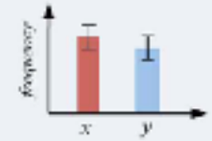
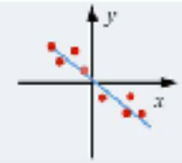
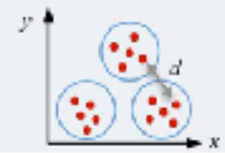
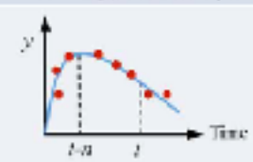
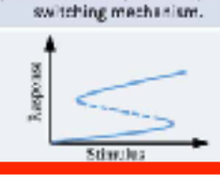
Oxford dictionary:

“A simplified description, especially a mathematical one, of a system or process, to assist calculations and predictions.”

A mathematical model is
"a **representation** of the **essential aspects** of a system … which presents **knowledge** of that system in **usable form**."

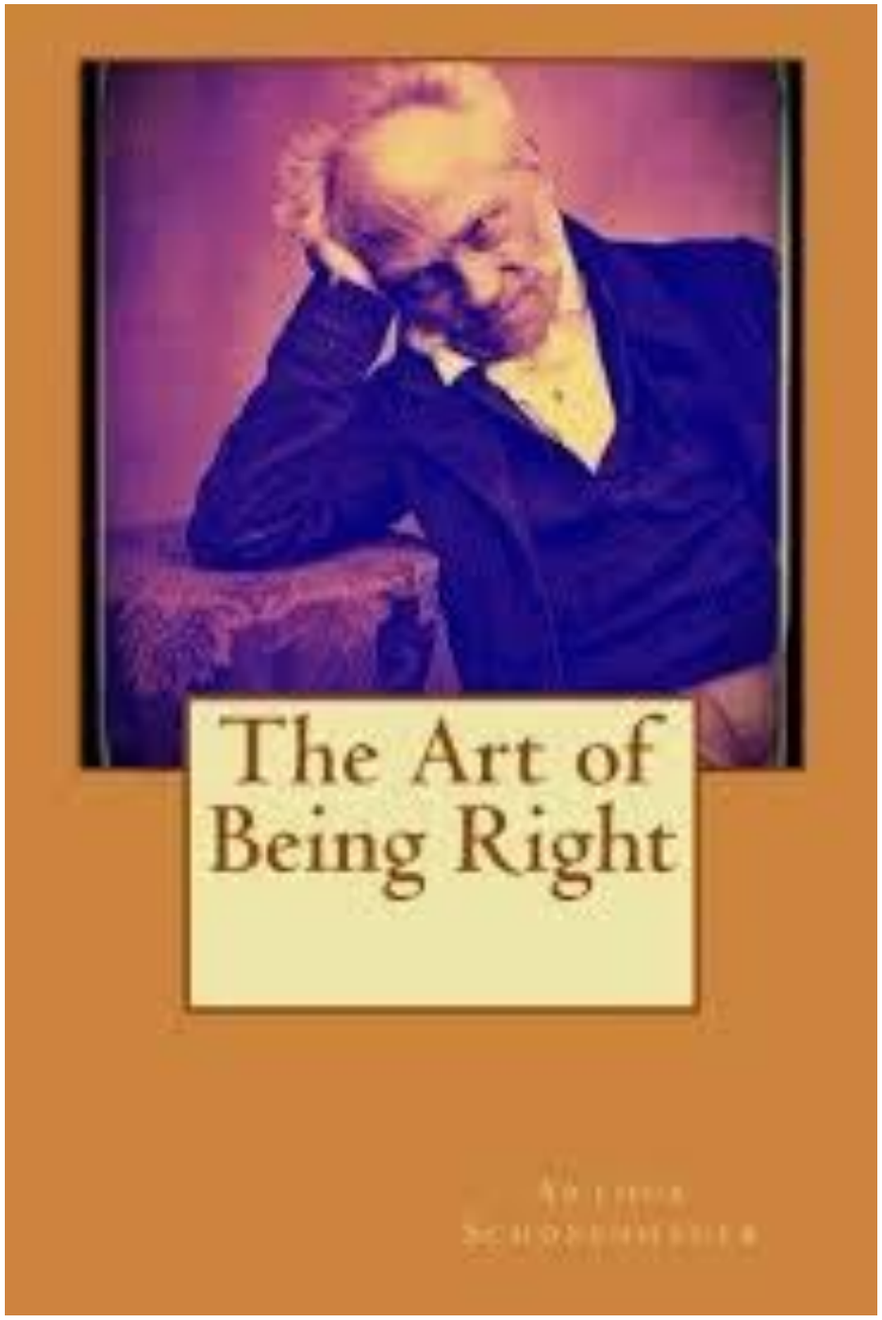
Eykhoff (1974)

Usable for what?

Testing for differences	Significance Analysis	Is there a difference between the two groups?	Ease of experiments
"before/after" "with/without"	$95\%\ \mathrm{CI} = \overline{X} \pm \left(t_{(n-1)}(5\%) \times \overline{\mathrm{SE}}\right)$		
Analysing covariaton	Correlation Analysis	Does a small/large value of x, coincide with a small/large value of y?	
"coincides with"	$r_{x,y} = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \sum y_i^2}}$		
Identifying groups	Cluster/Discriminant Analysis	Can the data be grouped on the basis of features x and y?	
"is similar to"	$D = \min_i \left\{ \min_j \left\{ \frac{d(i,j)}{\max_k d'(k)} \right\} \right\}$		
Condensing data	Component/Factor Analysis	Explain the variability through a weighted linear combination of principal components.	
"reduces to"	$X = U \Sigma W^T$		
Fitting data	Regression Analysis	What is the predicted value of y, given x_1 and x_2?	
"relates to"	$y = f(x_1, \ldots, x_n)$		
Numerical predictions	Time Series Analysis	What is the predicted value of y, knowing past values of y?	
"follows"	$y(t) = f(y(t-1), y(t-2), \ldots)$		
Analyzing influences	Bayesian (network) Analysis	From the observed evidence, H_2 is most probable.	Power of explanation
"Given E, the probability of H is"	$P(H_j \mid E) = \frac{\Pr(E \mid H_j)\Pr(H_j)}{\sum \Pr(E \mid H_j)\Pr(H_j)}$		
Investigating mechanisms	Dynamical Systems Theory	The response can be explained by a (bistable) switching mechanism.	
"causally entails" "If ... Then ..."	$\frac{d}{dt} x_i = f_i(x_1(t), \ldots, x_n(t), u(t))$		

Wolkenhauer, *Frontier in Physiology*, 2014.

Formal modelling is crucial for decision making



Arthur Schopenhauer

The Art of Being Right:
38 Ways to Win an Argument

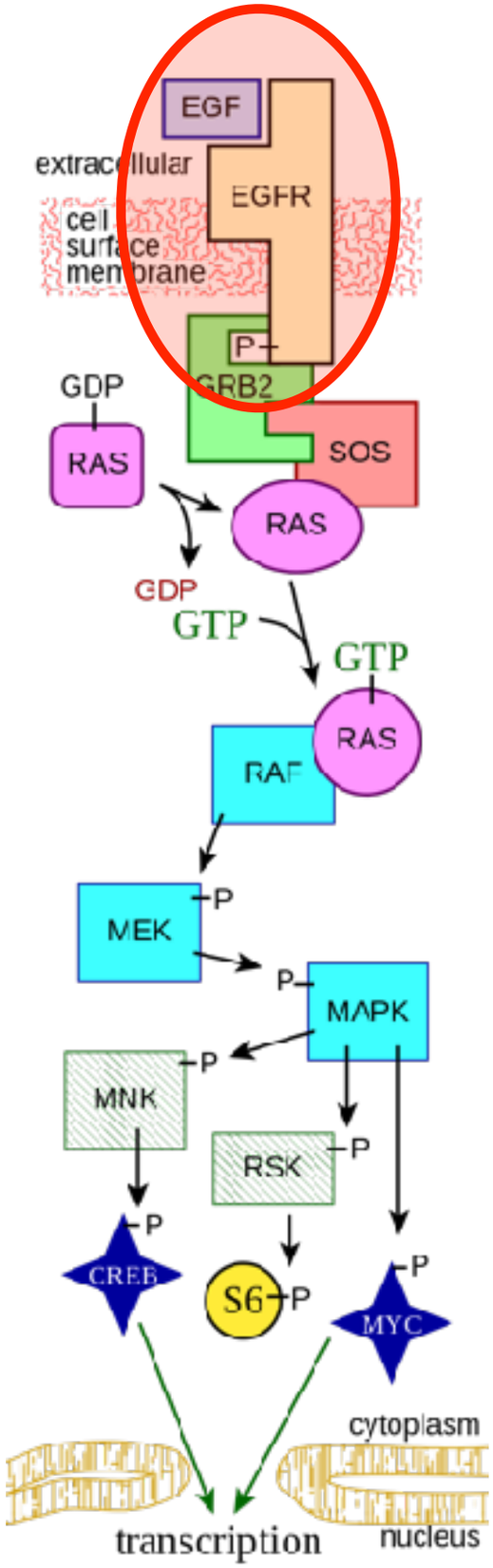
(German: Eristische Dialektik: **Die Kunst, Recht zu behalten**; 1831) is an acidulous and sarcastic treatise written by the German philosopher Arthur Schopenhauer in sarcastic deadpan.

Overview

- How do we derive the models?
- How can a parameter estimation problem be formulated?
- How can sensitivities and gradients be computed?
- How does optimisation work?
- How can we deal with relative data?
- How can we deal with outliers?
- How well are the parameters determined?
- What tools are available?
- A large-scale model of cancer signalling

How do we derive the models?

Formulation of models



Reaction:

$$\mathrm{EGF} + \mathrm{EGFR} \xrightarrow{k_1} \mathrm{EGF{:}EGFR}$$

Mathematical description:

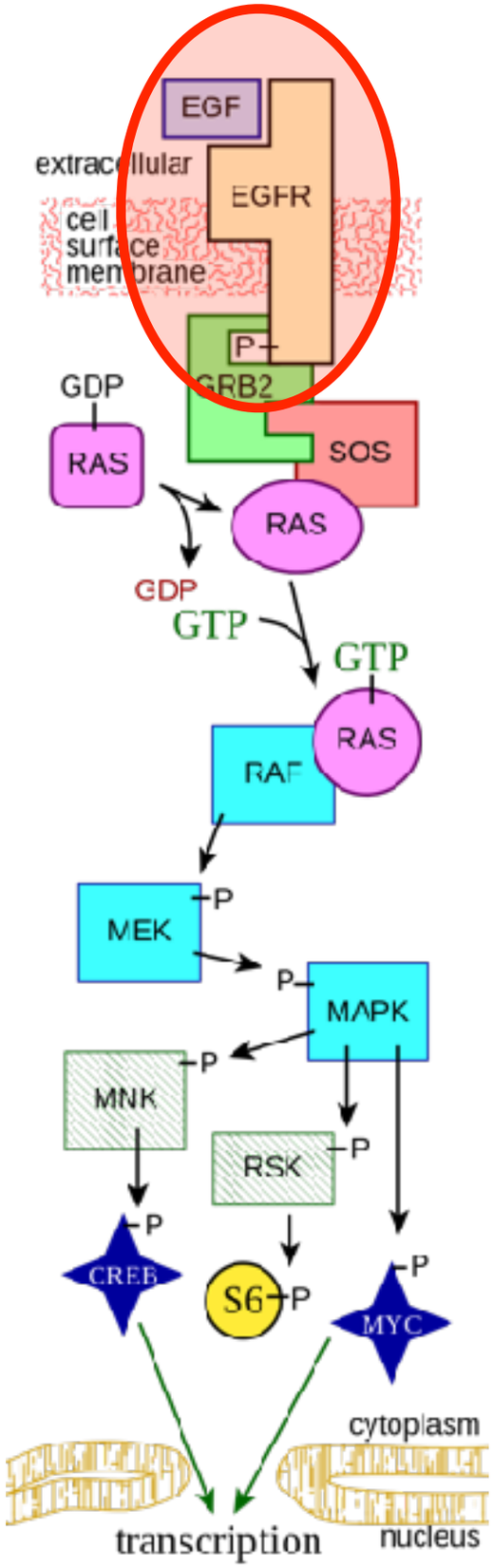
reaction stoichiometry

$$\begin{array}{c} \mathrm{EGF} \\ \mathrm{EGFR} \\ \mathrm{EGF{:}EGFR} \\ \vdots \end{array} \quad \begin{pmatrix} -1 \\ -1 \\ +1 \\ \vdots \end{pmatrix}$$

reaction flux

$$v_1 = k_1[\mathrm{EGF}][\mathrm{EGFR}]$$

Formulation of models



Reaction:

$$\mathrm{EGF} + \mathrm{EGFR} \xrightarrow{k_1} \mathrm{EGF{:}EGFR}$$

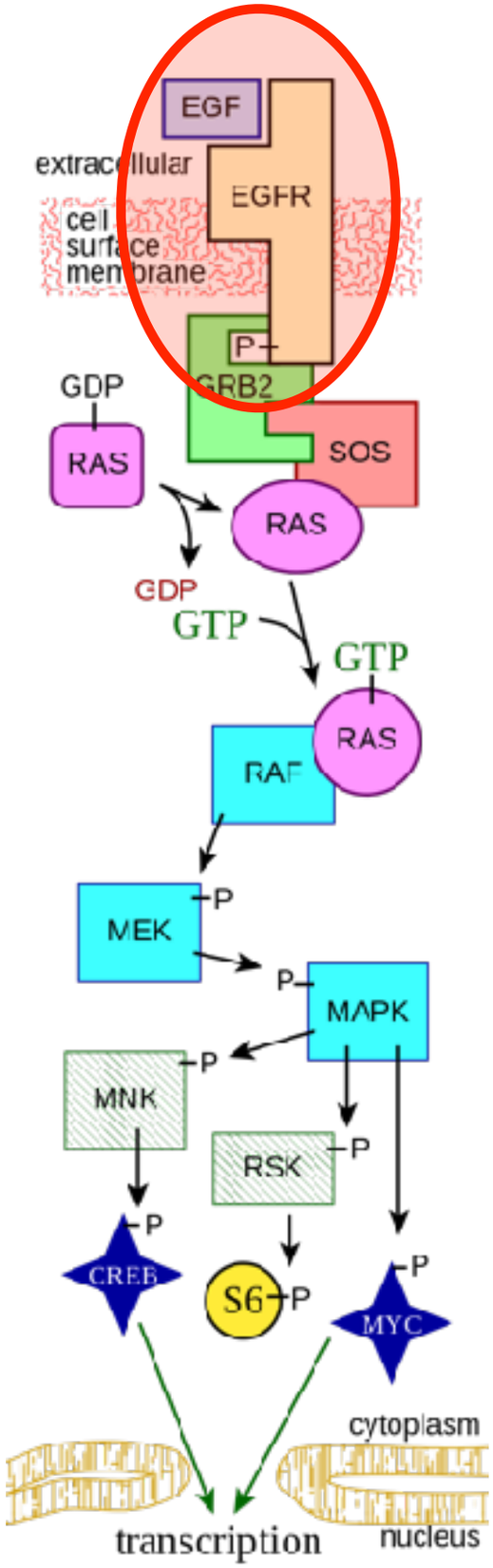
Mathematical description:

reaction stoichiometry *reaction flux*

$$[\mathrm{EGF}](t+\Delta t) = [\mathrm{EGF}](t) \; -1 \; \cdot k_1[\mathrm{EGF}][\mathrm{EGFR}] \cdot \Delta t$$

new concentration *old concentration* *change*

Formulation of models



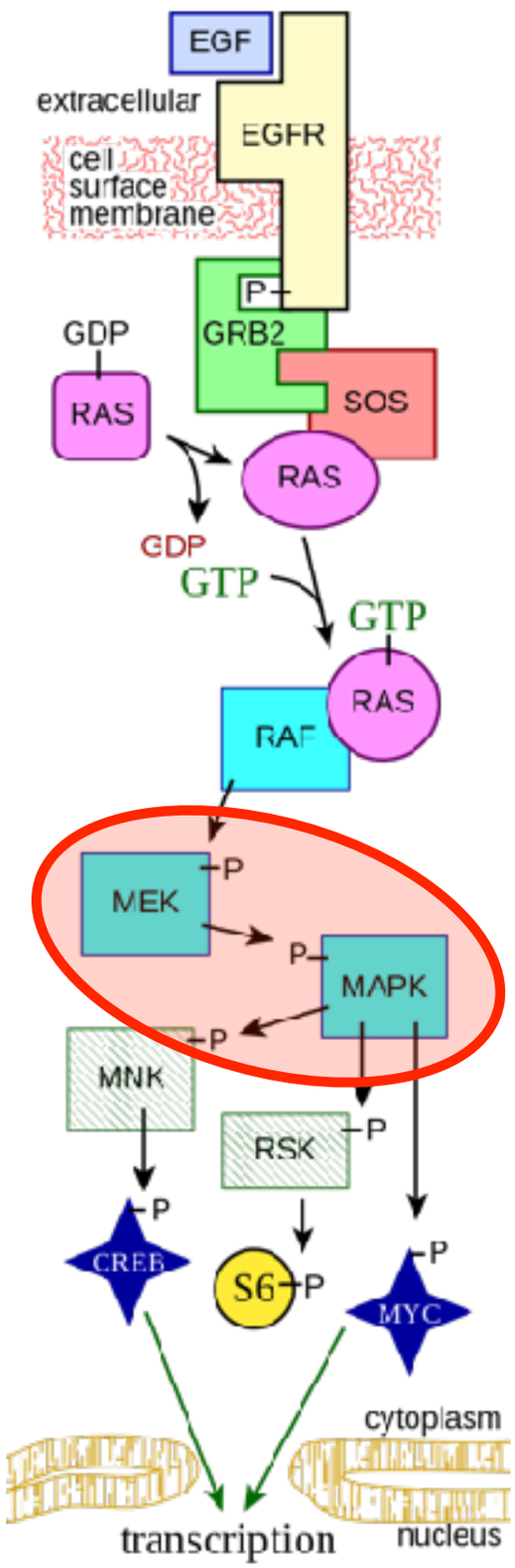
Reaction:

$$\mathrm{EGF} + \mathrm{EGFR} \xrightarrow{k_1} \mathrm{EGF{:}EGFR}$$

Mathematical description:

$$\frac{d[\mathrm{EGF}]}{dt} = -k_1[\mathrm{EGF}][\mathrm{EGFR}] + \ldots$$

Formulation of models



Reaction:

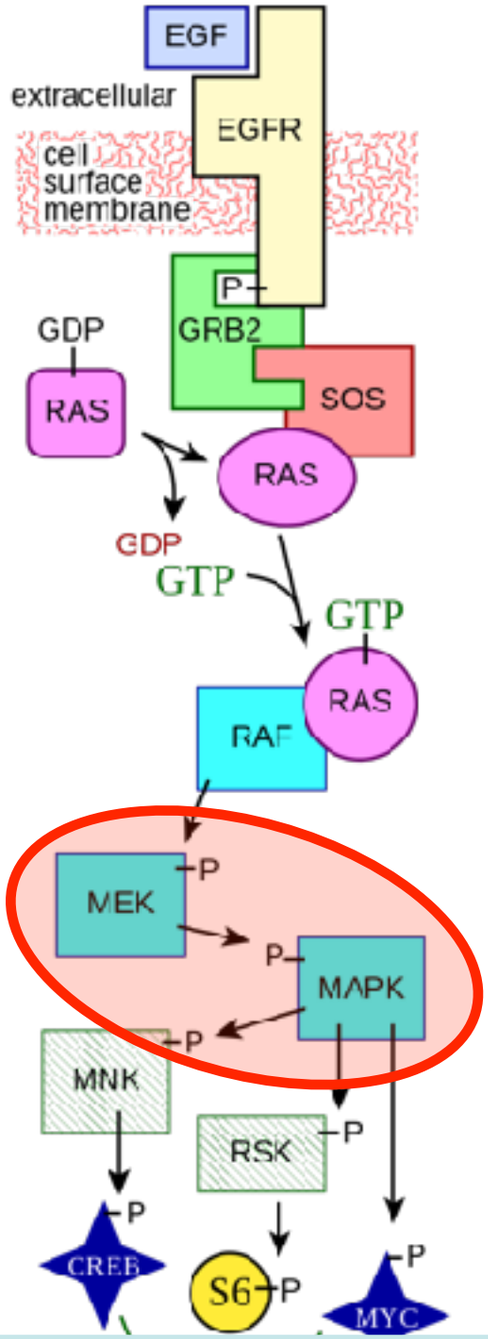
$$\mathrm{MAPK} + \mathrm{pMEK} \xrightarrow{k_8} \mathrm{MAPK{:}pMEK}$$

Mathematical description:

reaction stoichiometry *reaction flux*

$$\begin{array}{c} \vdots \\ \mathrm{pMEK} \\ \mathrm{MAPK} \\ \mathrm{MAPK{:}pMEK} \\ \vdots \end{array} \quad \begin{pmatrix} \vdots \\ -1 \\ -1 \\ +1 \\ \vdots \end{pmatrix} \qquad v = k_8[\mathrm{MAPK}][\mathrm{pMEK}]$$

Formulation of models



Reaction:

$$\mathrm{MAPK} + \mathrm{pMEK} \xrightarrow{k_8} \mathrm{MAPK{:}pMEK}$$

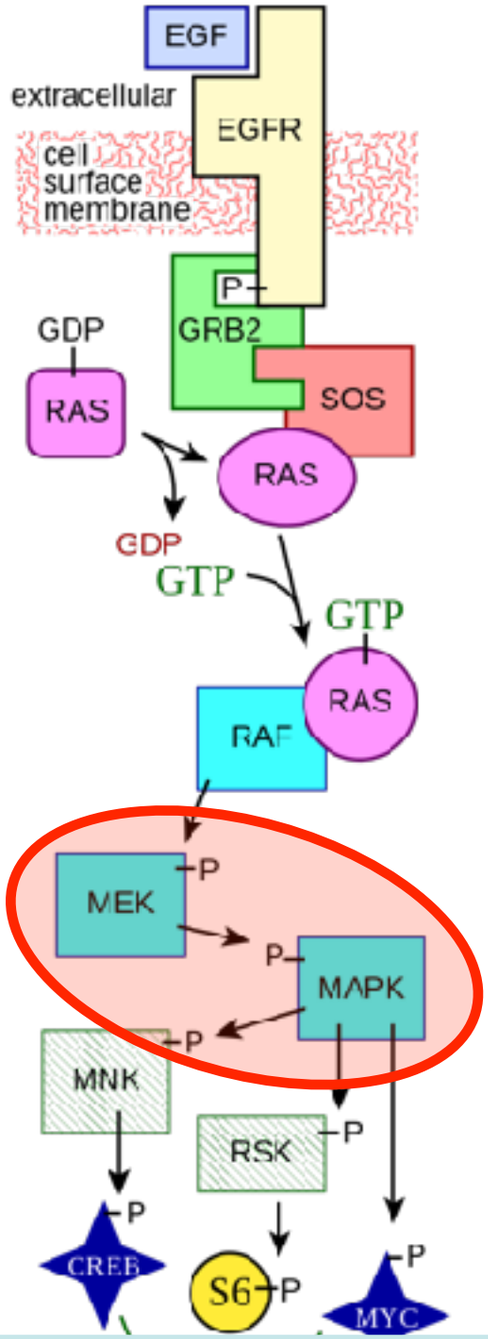
Mathematical description:

reaction stoichiometry *reaction flux*

$$\begin{array}{c} \vdots \\ \mathrm{pMEK} \\ \mathrm{MAPK} \\ \mathrm{MAPK{:}pMEK} \end{array} \quad \begin{pmatrix} \vdots \\ -1 \\ -1 \\ +1 \end{pmatrix} \qquad v = k_8[\mathrm{MAPK}][\mathrm{pMEK}]$$

We get a system of equations describing the biochemistry.

Formulation of models



Reaction:

$$\mathrm{MAPK} + \mathrm{pMEK} \xrightarrow{k_8} \mathrm{MAPK{:}pMEK}$$

Mathematical description:

reaction stoichiometry *reaction flux*

$$\begin{array}{c} \vdots \\ \mathrm{pMEK} \\ \mathrm{MAPK} \\ \mathrm{MAPK{:}pMEK} \end{array} \begin{pmatrix} \vdots \\ -1 \\ -1 \\ +1 \end{pmatrix} \qquad v = k_8[\mathrm{MAPK}][\mathrm{pMEK}]$$

Exchange und reuse is simplified by community standards such as SBML.

Formulating mechanistic models is not difficult

Article

TRANSPARENT PROCESS OPEN ACCESS

molecular systems biology

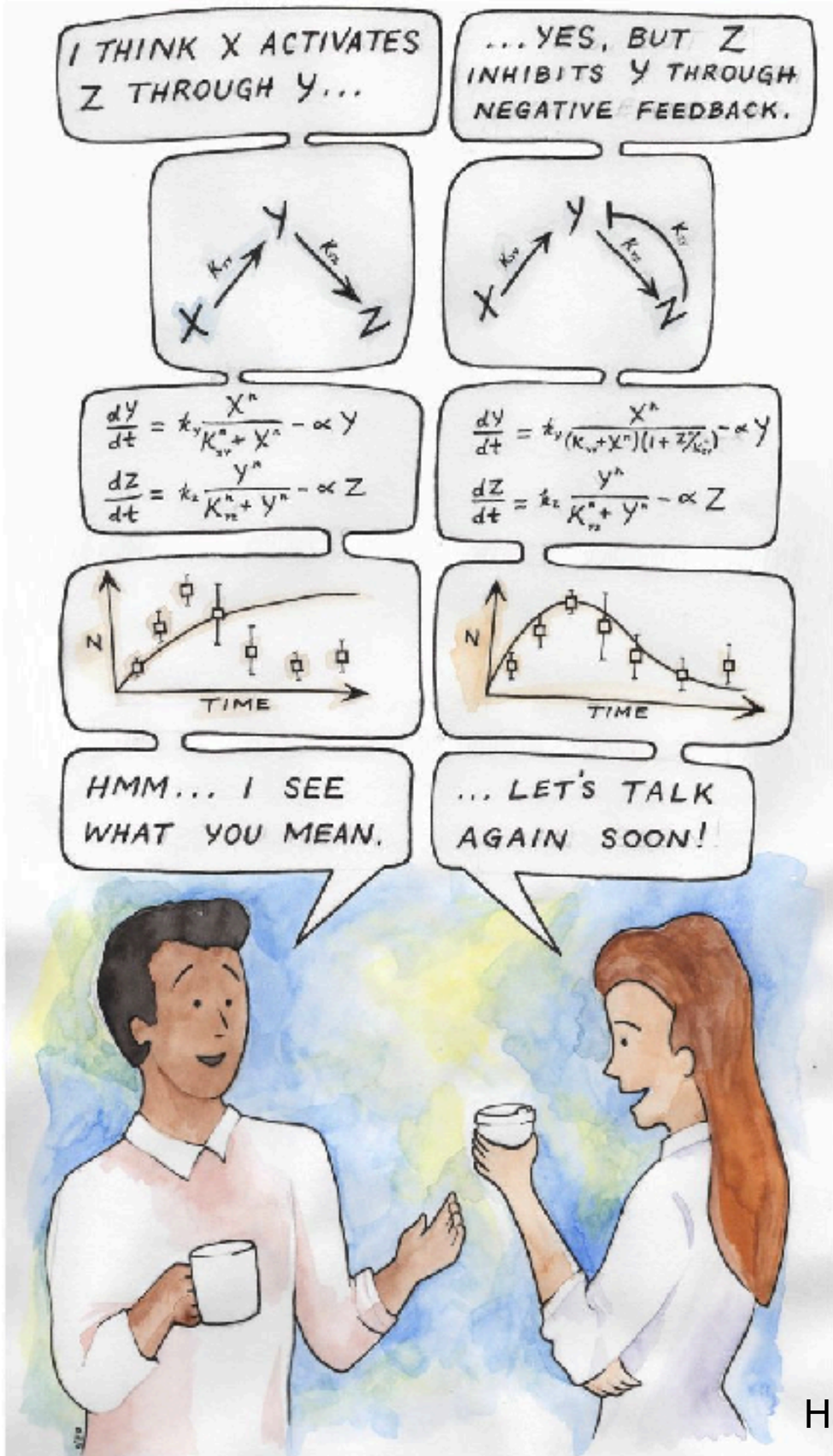
From word models to executable models of signaling networks using automated assembly

Benjamin M Gyori[1,†], John A Bachman[1,†], Kartik Subramanian[1], Jeremy L Muhlich[1], Lucian Galescu[2] & Peter K Sorger[1,*]

Abstract

Word models (natural language descriptions of molecular mechanisms) are a common currency in spoken and written communication in biomedicine but are of limited use in predicting the behavior of complex biological networks. We present an approach to building computational models directly from natural language using automated assembly. Molecular mechanisms described in simple English are read by natural language processing algorithms, converted into an intermediate representation, and assembled into executable or network models. We have implemented this approach in the Integrated Network and Dynamical Reasoning Assembler (INDRA), which draws on existing natural language processing systems as well as pathway information in Pathway

in reaction diagrams (familiar graphs involving forward and reverse arrows) and analyzed algebraically. As such systems became more complex and grew to include large networks in mammalian cells, word models (natural language descriptions) became the dominant way of describing biochemical processes; word models are frequently illustrated using pictograms and informal schematics. However, formal approaches are generally required to understand dynamics, multi-component switches, bistability, etc. Dynamical models and systems theory have proven extremely effective in elucidating mechanisms of all-or-none response to apoptosis-inducing ligands (Rehm *et al*, 2002; Albeck *et al*, 2008), sequential execution of cell cycle phases (Chen *et al*, 2004), the interplay between stochastic and deterministic reactions in the control of cell fate following DNA damage (Purvis *et al*, 2012), drug sensitivity



Haggerty & Purvis, MSB, 2017

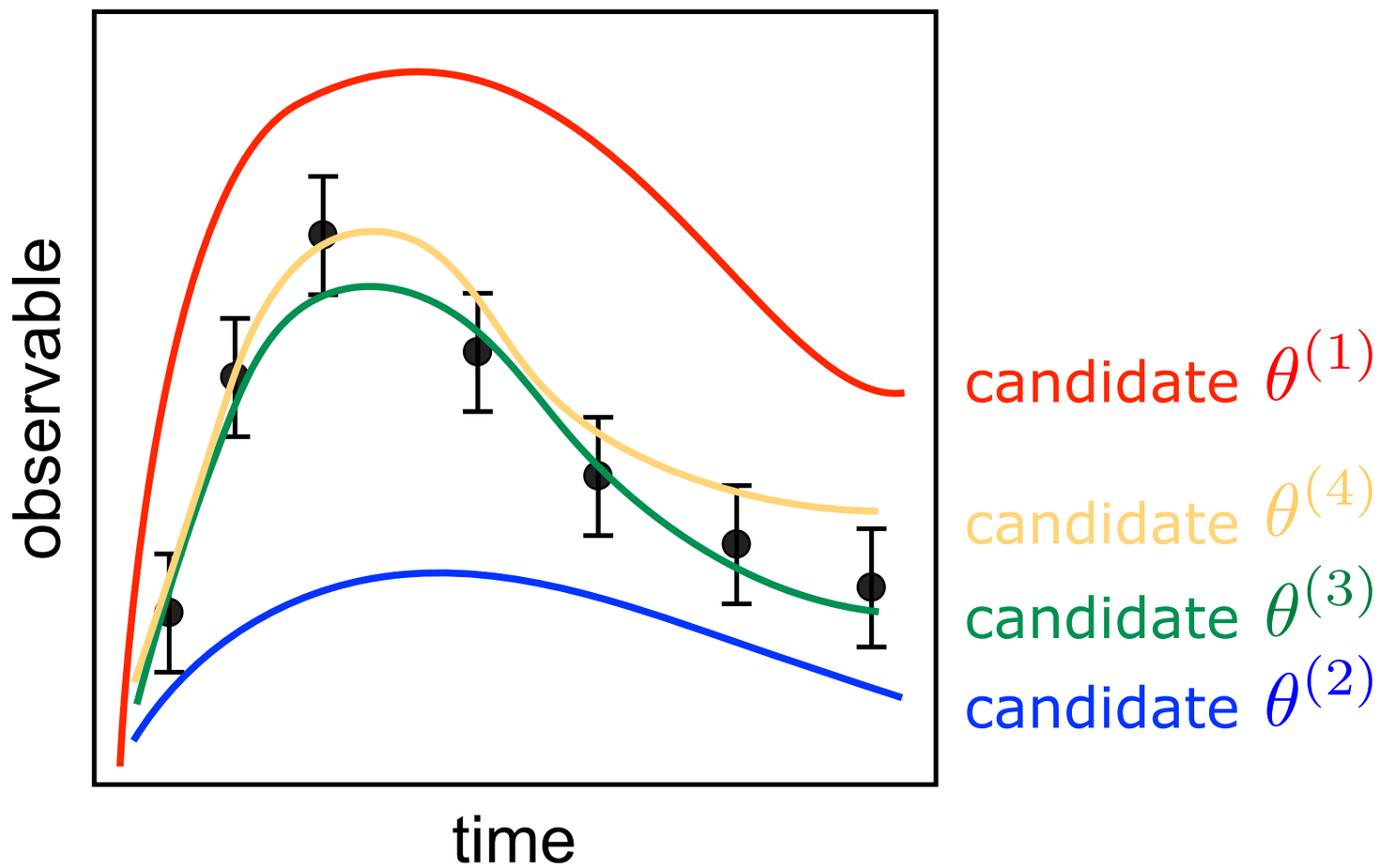
How can a parameter estimation problem be formulated?

Basic idea of parameter estimation

ODE model:

compounds: $\dot{x} = f(x, \theta, u(t)), \qquad x(0) = x_0(\theta, u)$

observables: $y(t_i, \theta) = h(x(t_i), \theta)$



How to distinguish quantitatively?

Formulation of estimation problem for ODEs

ODE model:

compounds: $\dot{x} = f(x, \theta, u(t)),$ $\quad x(0) = x_0(\theta, u)$

observables: $y(t_i, \theta) = h(x(t_i), \theta)$

Measurement: $\bar{y}(t_i) = y(t_i) + \epsilon(t_i),$ $\quad \epsilon_k(t_i) \sim \mathcal{N}(0, \sigma_{k,i}^2)$

Likelihood function:

$$p(\mathcal{D}|\theta) = \prod_i \prod_k \frac{1}{\sqrt{2\pi}\sigma_{k,i}} \exp\left\{ -\frac{1}{2} \left(\frac{\bar{y}_k(t_i) - y_k(t_i, \theta)}{\sigma_{k,i}} \right)^2 \right\}$$

Maximum likelihood estimation:

optimisation of the likelihood function $\quad \max_\theta p(\mathcal{D}|\theta)$

Relation: Maximum Likelihood $\Leftrightarrow$ Least squares

Negative log-likelihood function for independent, additive normally distributed measurement noise:

$$p(\mathcal{D}|\theta) = \prod_k \prod_j \frac{1}{\sqrt{2\pi}\sigma_{jk}(\theta)} \exp\left\{ -\frac{1}{2} \left(\frac{\bar{y}_{jk} - y_j(t_k, \theta)}{\sigma_{jk}(\theta)} \right)^2 \right\}$$

$$\Rightarrow J(\theta) = -\log p(\mathcal{D}|\theta) = \frac{1}{2} \sum_k \sum_j \log(2\pi\sigma_{jk}^2(\theta)) + \left(\frac{\bar{y}_{jk} - y_j(t_k, \theta)}{\sigma_{jk}(\theta)} \right)^2$$

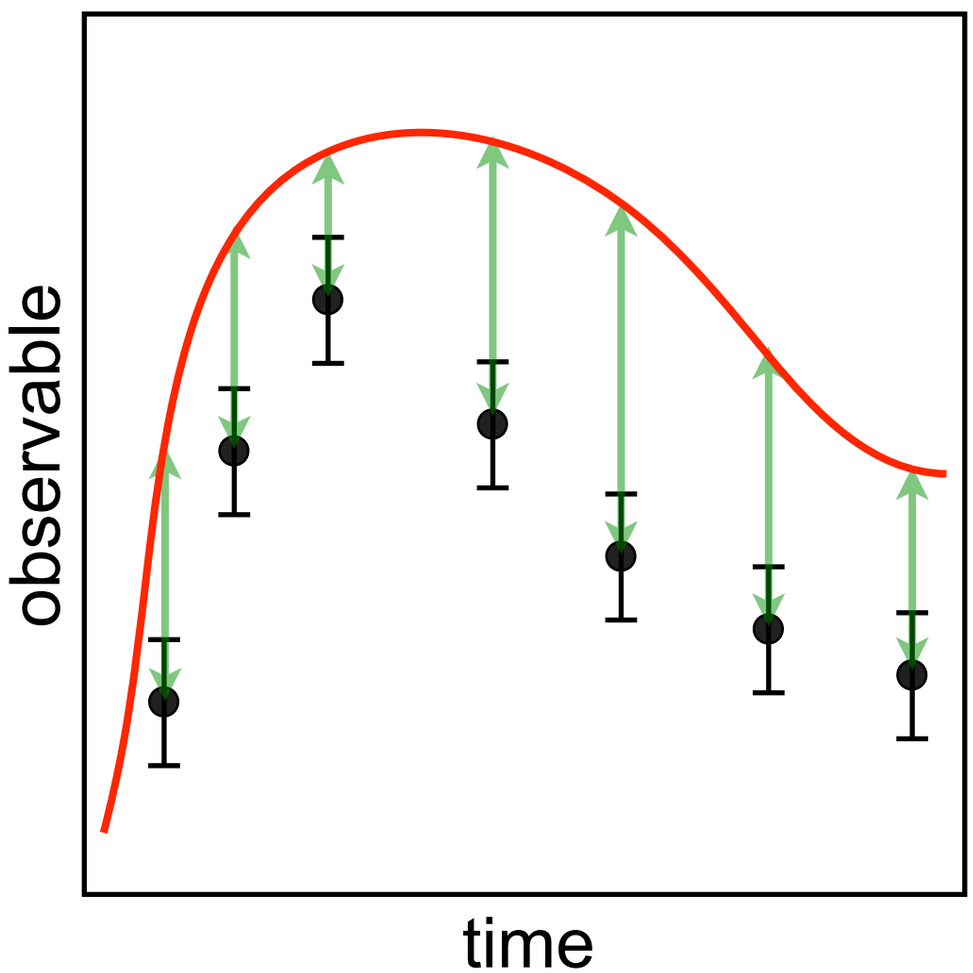
For known noise variance $\sigma_{jk}^2(\theta) \neq \mathrm{fnc}(\theta)$, the term $\log(2\pi\sigma_{jk}^2(\theta))$ is constant and the objective function can be simplified to

$$J(\theta) = \frac{1}{2} \sum_k \sum_j \left(\frac{\bar{y}_{jk} - y_j(t_k, \theta)}{\sigma_{jk}} \right)^2 .$$

$\Rightarrow$ objective function of the **weighted least squares estimator**

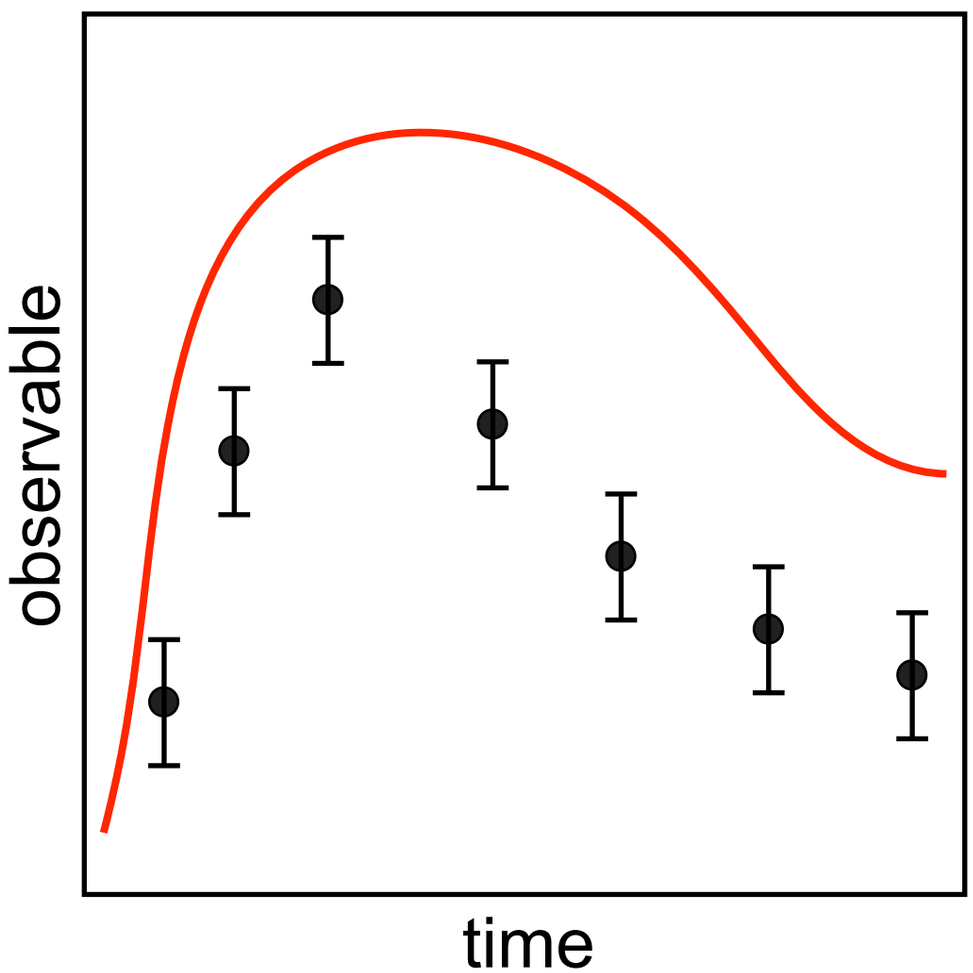
Evaluation of sum-of-squares objective function

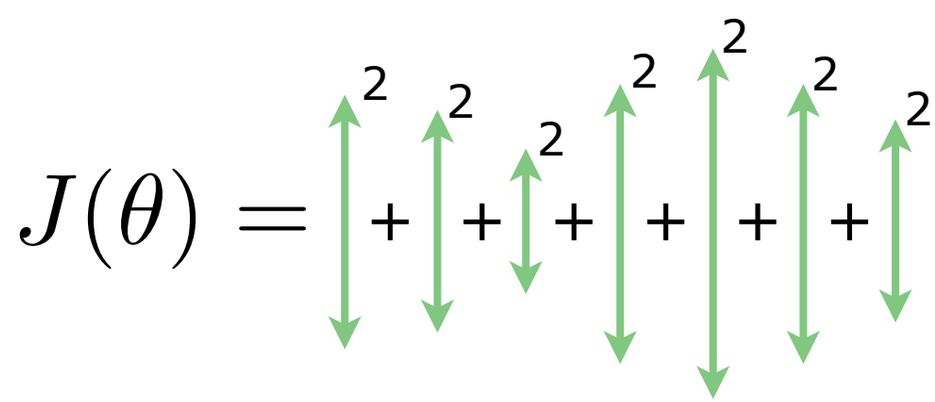
1. Simulation of mechanistic model
2. Calculation of distance between simulation results and data
3. Evaluation of sum-of-squares



Evaluation of sum-of-squares objective function

1. Simulation of mechanistic model
2. Calculation of distance between simulation results and data
3. Evaluation of sum-of-squares

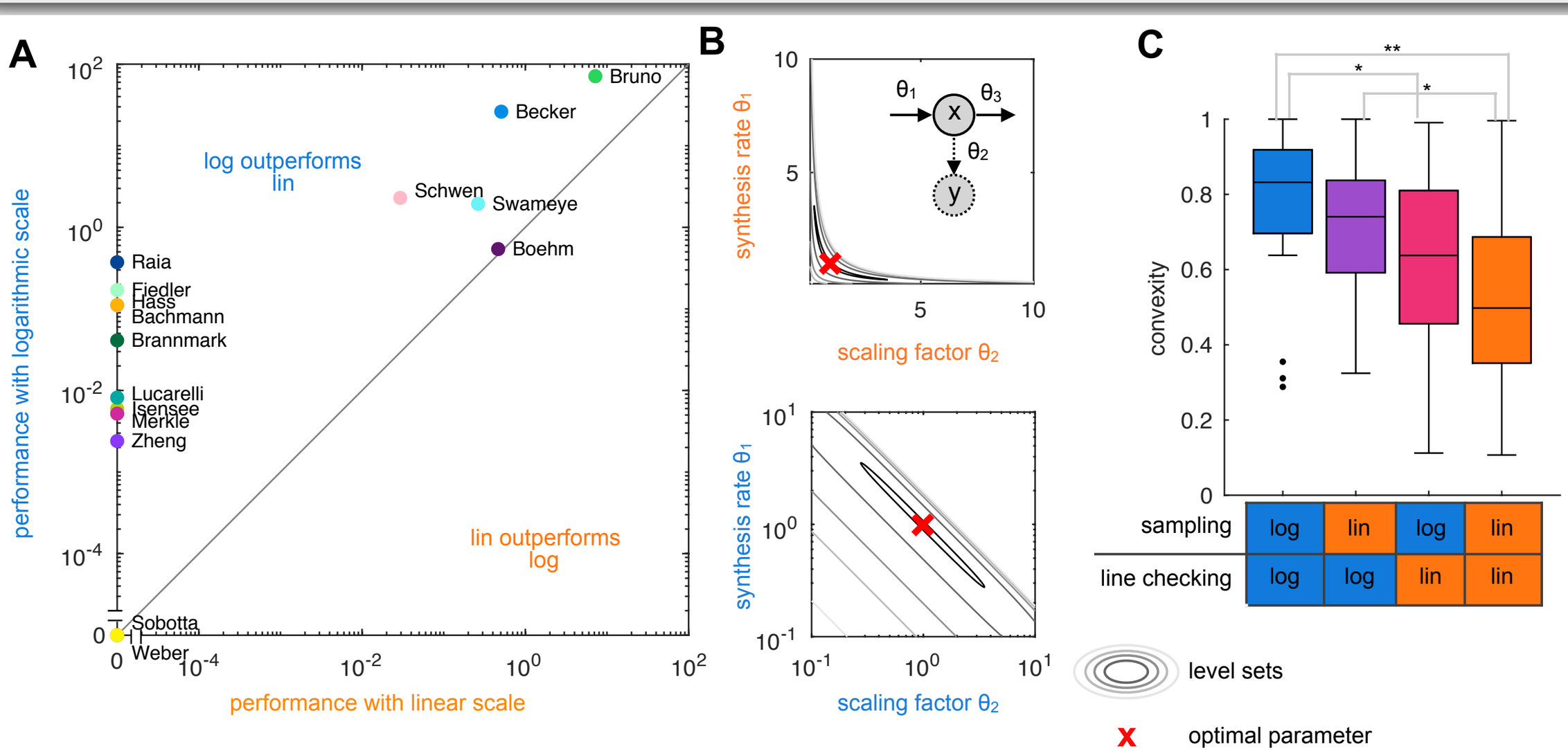




Log-transformation of parameters

For positive parameters $\theta \geq 0$, the ML estimate $\theta^{\mathrm{ml}} = \exp(\xi^{\mathrm{ml}})$ can be obtained by optimising the log-transformed parameter $\xi = \log(\theta)$,

$$\xi^{\mathrm{ml}} = \arg \min_{\xi \in \log(\Omega)} J(\exp(\xi)), \quad \text{subject to } \mathcal{M}(\exp(\xi)).$$

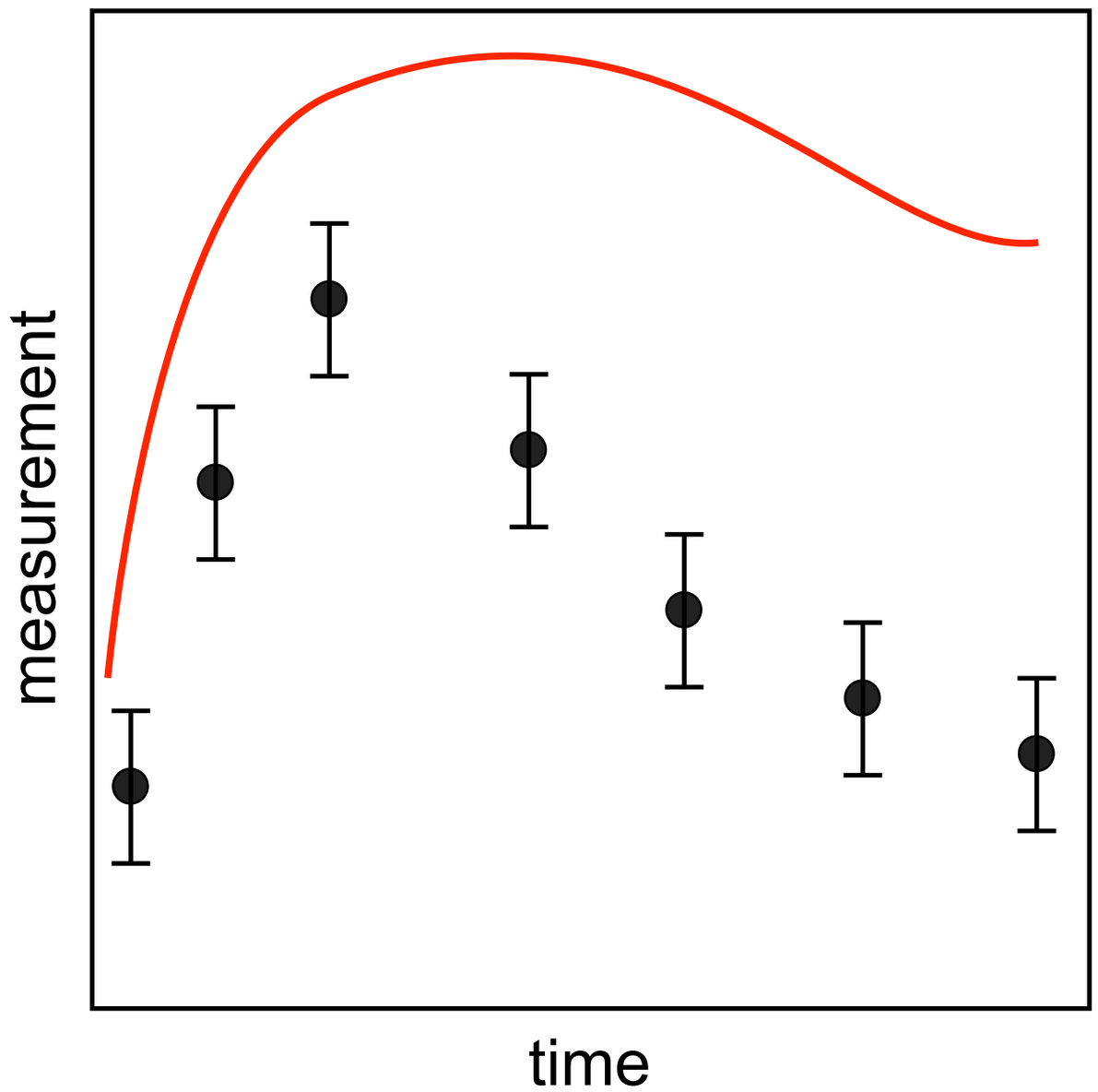


Hass et al. Benchmark Problems for Dynamic Modeling of Intracellular Processes. *Bioinformatics*, 2019.

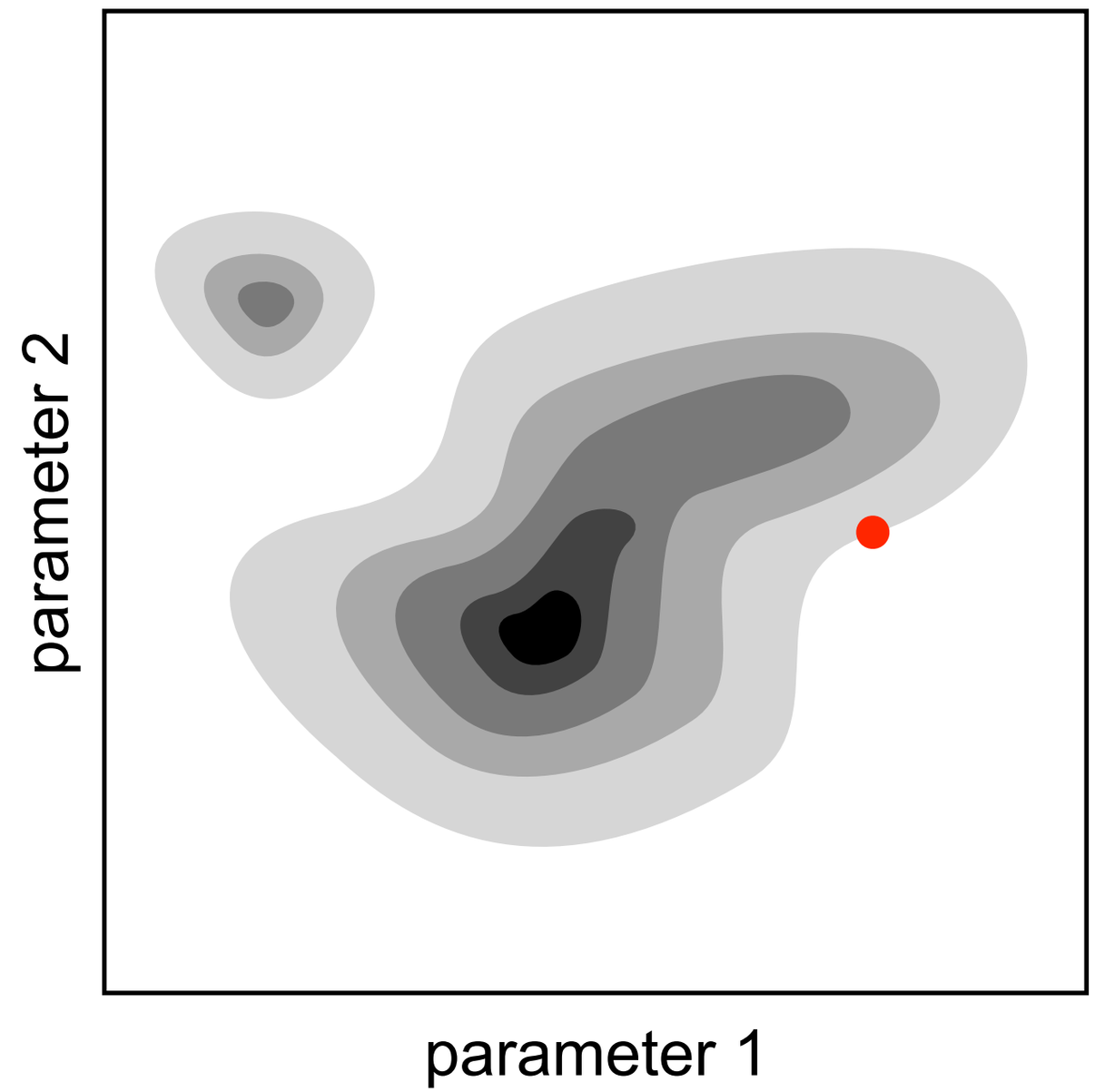
Sequential local optimisation

Goal: Construct a sequence of points along which the fit improves, meaning that the objective functions decreases

Model-data comparison



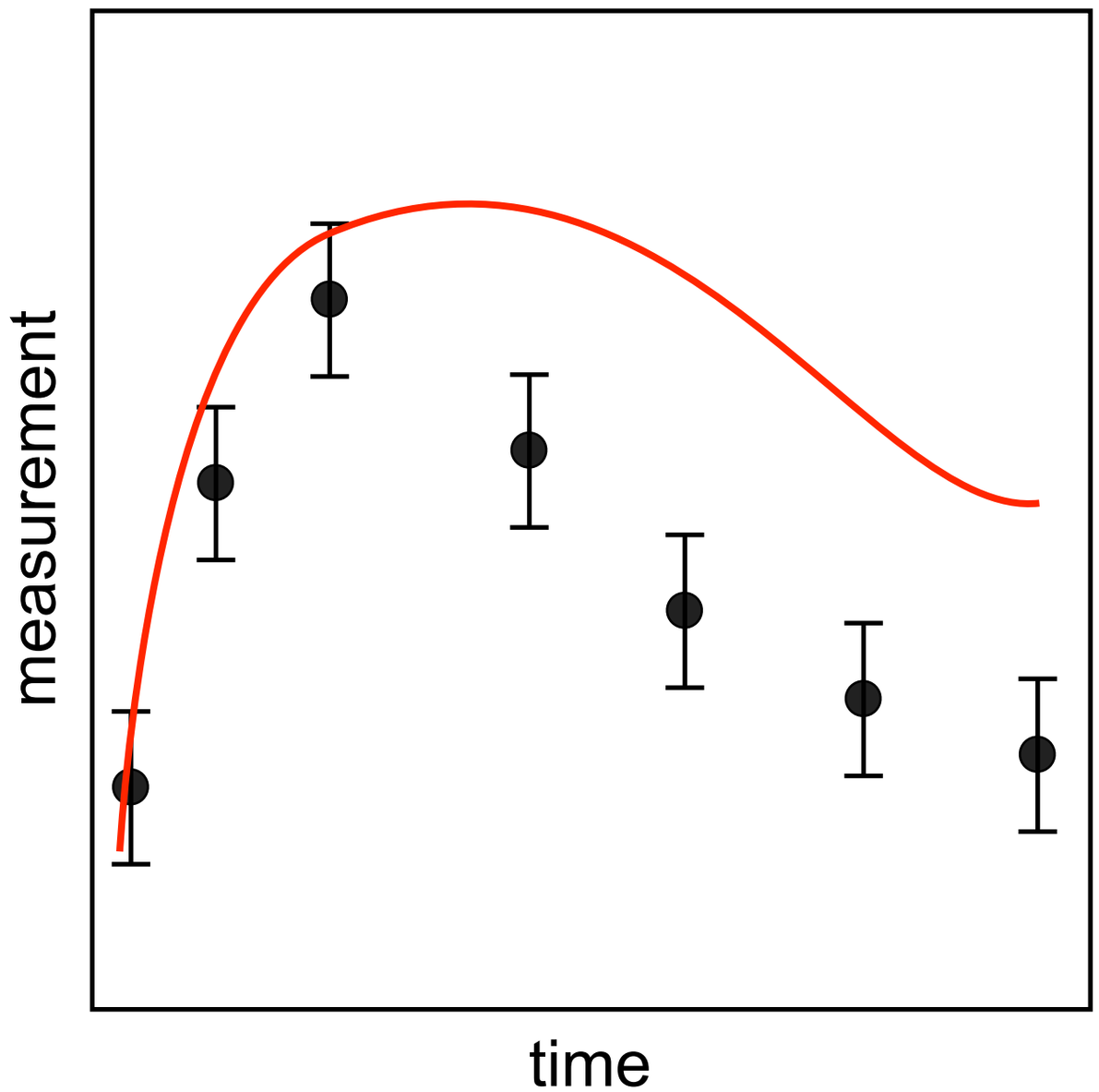
Objective function landscape



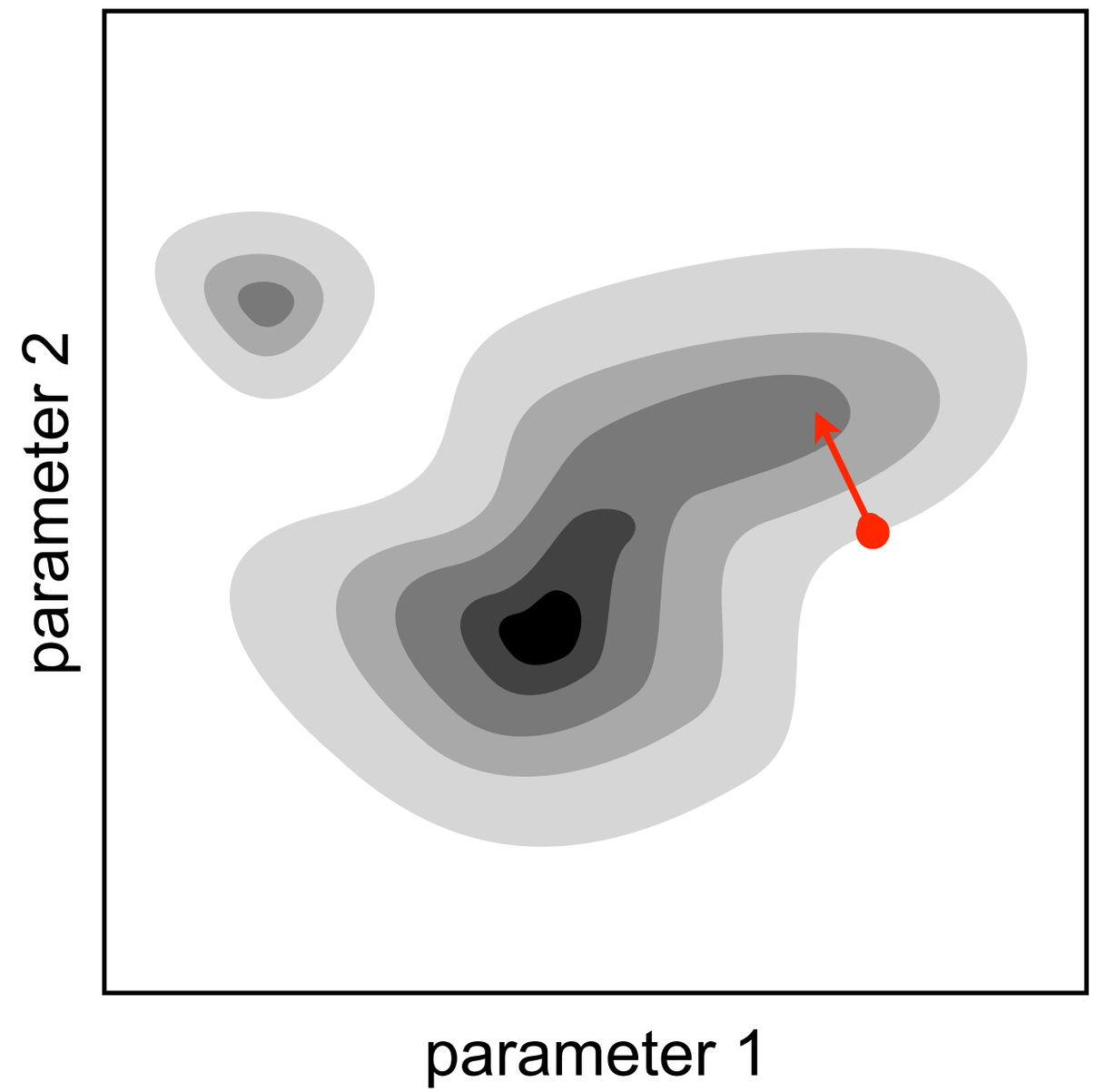
Sequential local optimisation

Goal: Construct a sequence of points along which the fit improves, meaning that the objective functions decreases

Model-data comparison



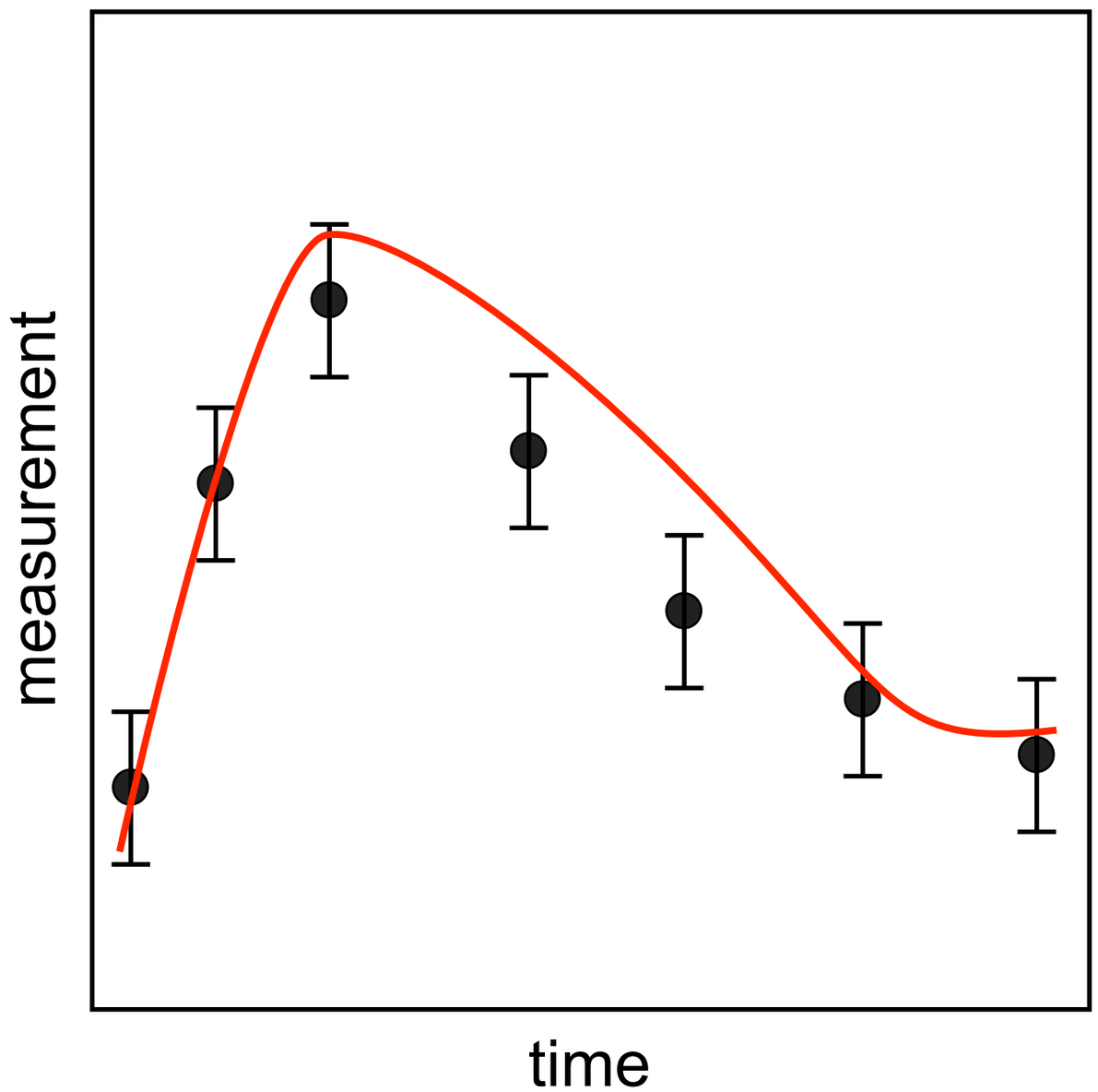
Objective function landscape



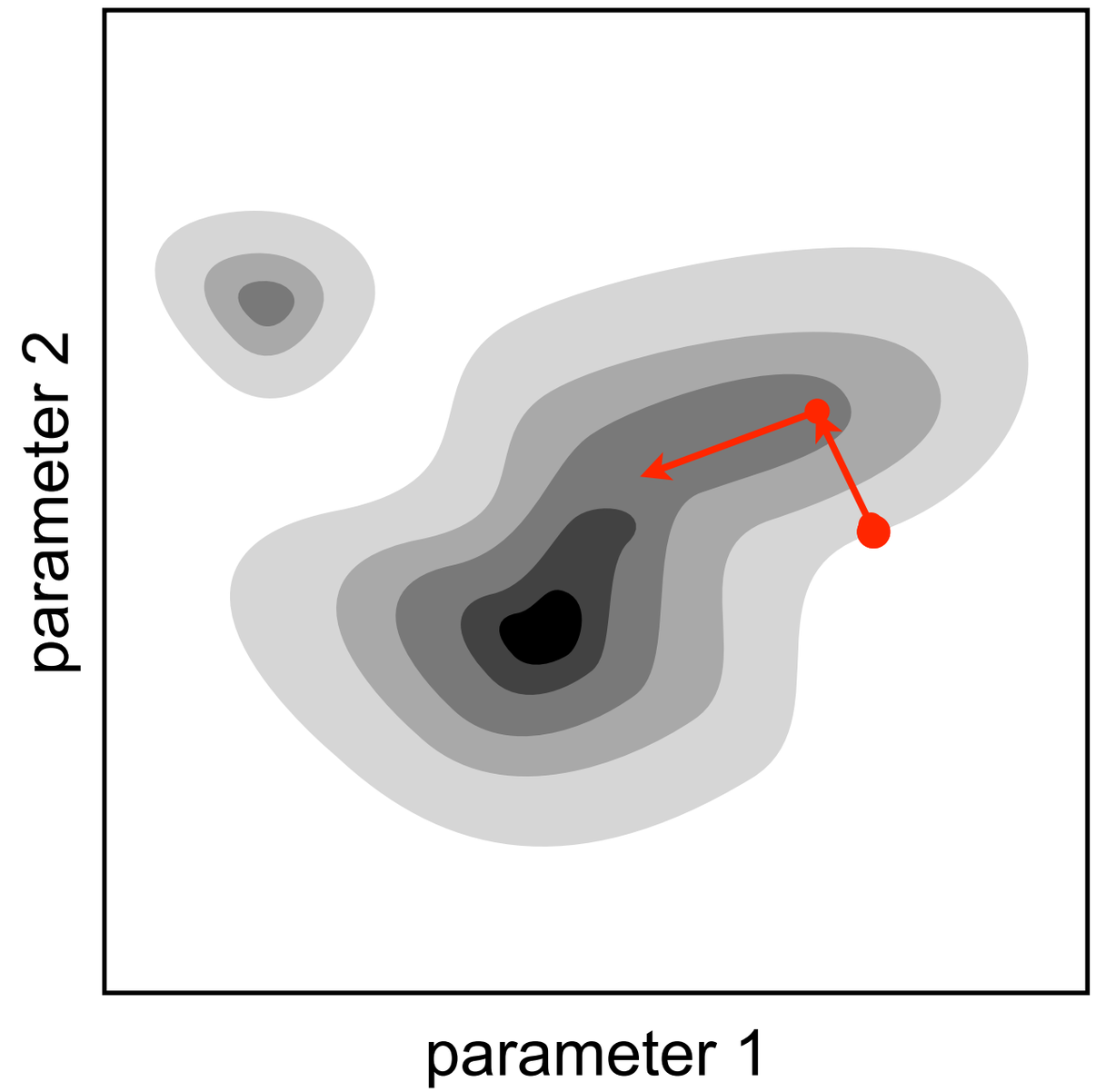
Sequential local optimisation

Goal: Construct a sequence of points along which the fit improves, meaning that the objective functions decreases

Model-data comparison



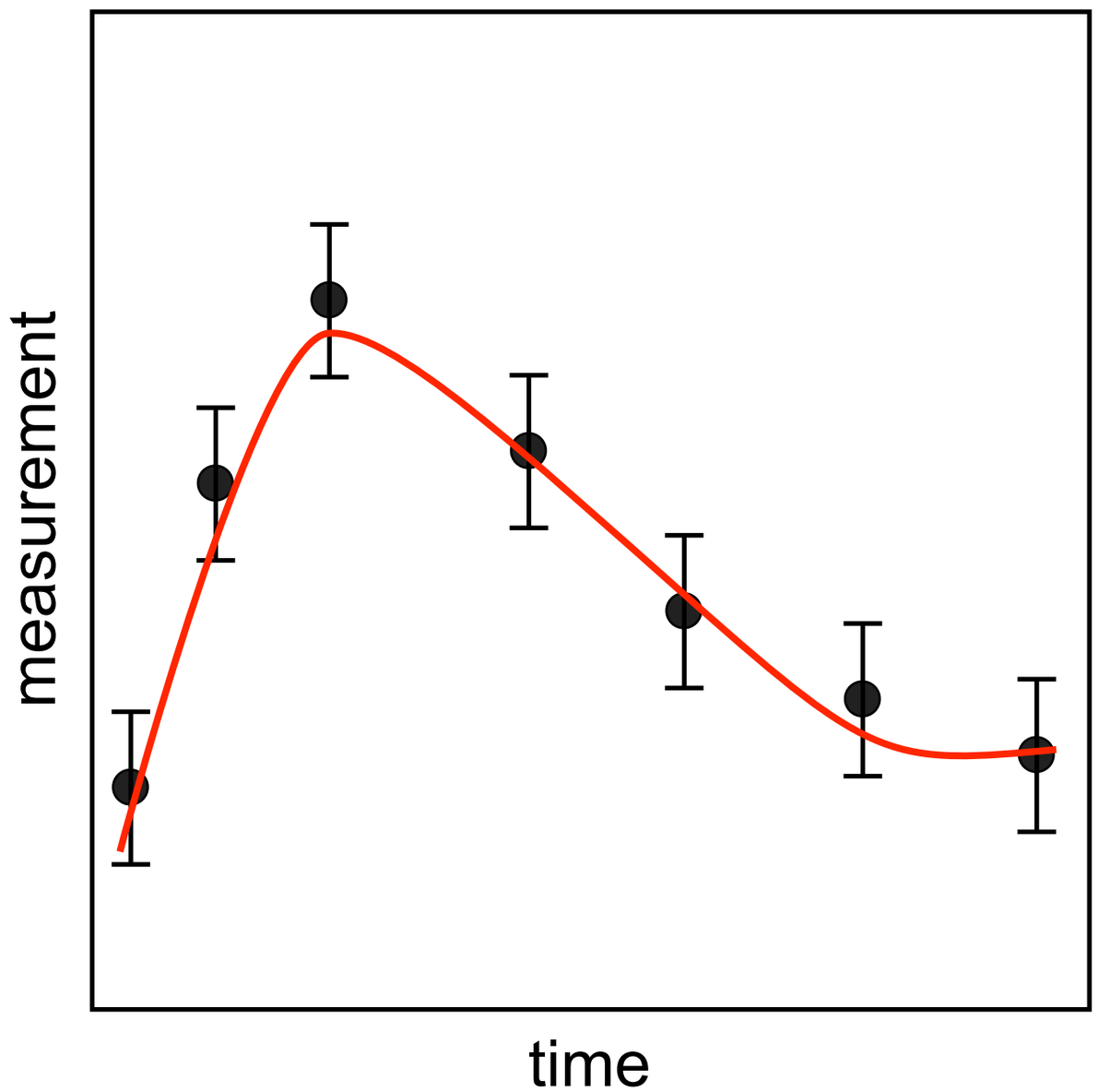
Objective function landscape



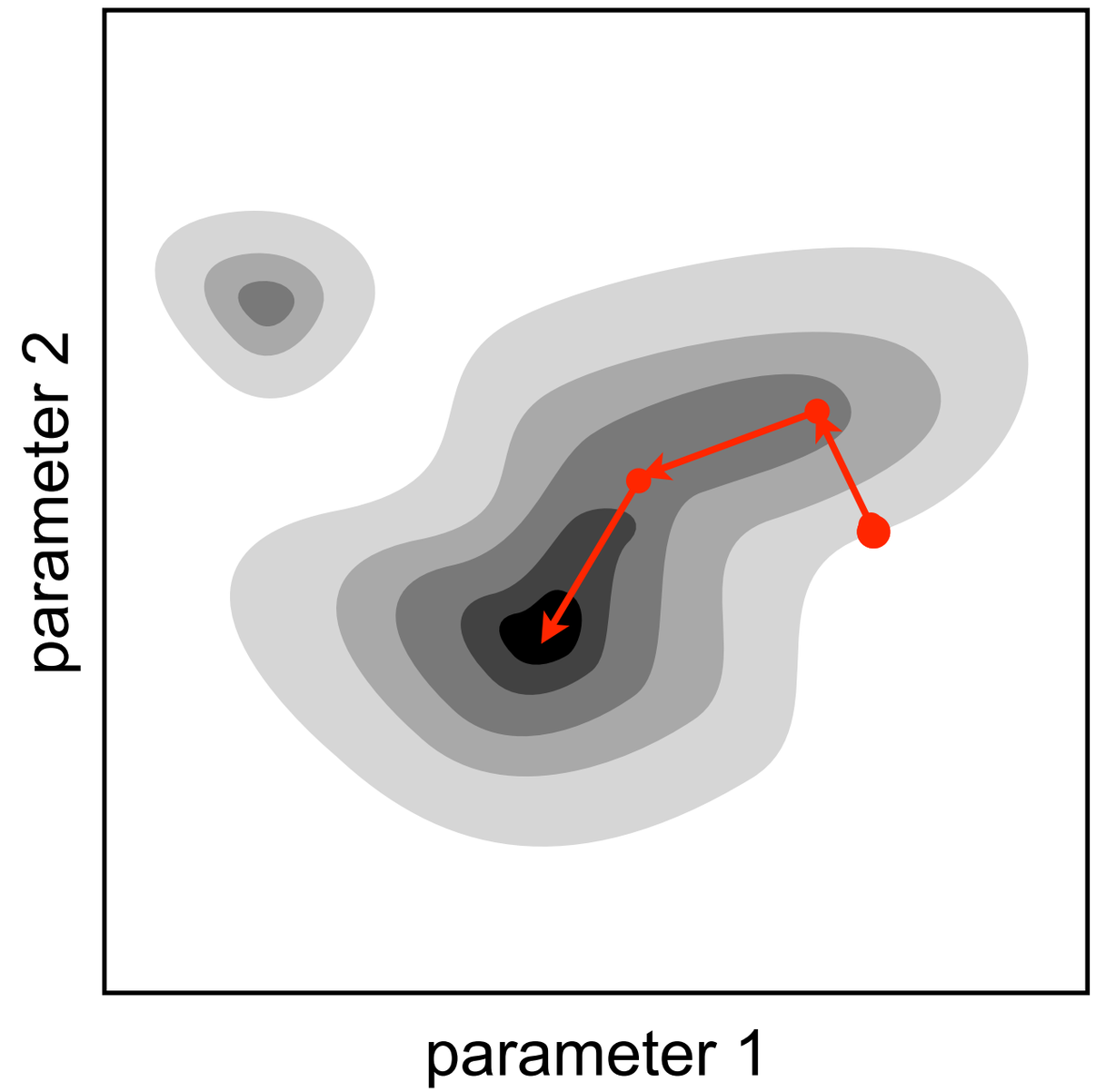
Sequential local optimisation

Goal: Construct a sequence of points along which the fit improves, meaning that the objective functions decreases

Model-data comparison



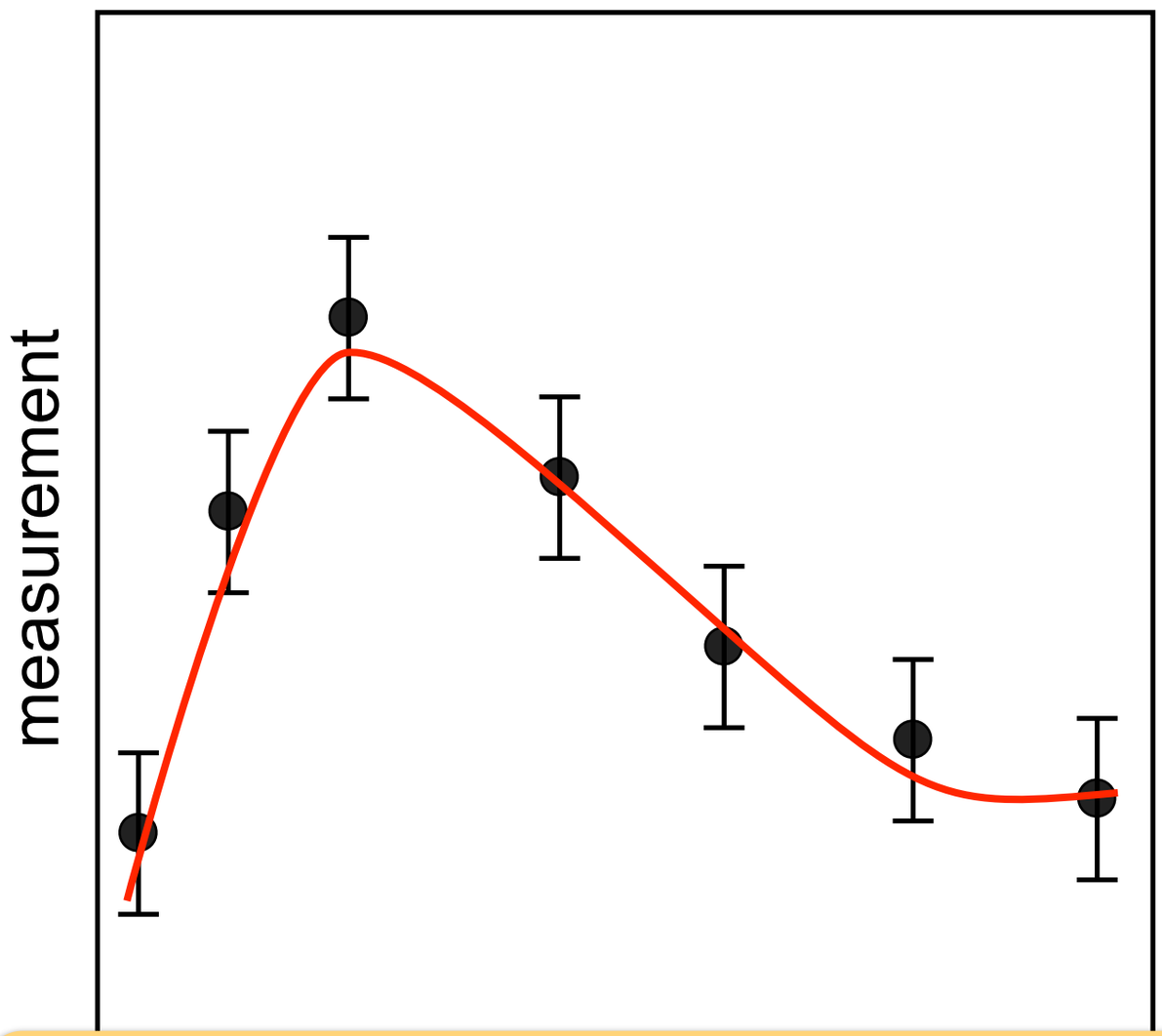
Objective function landscape



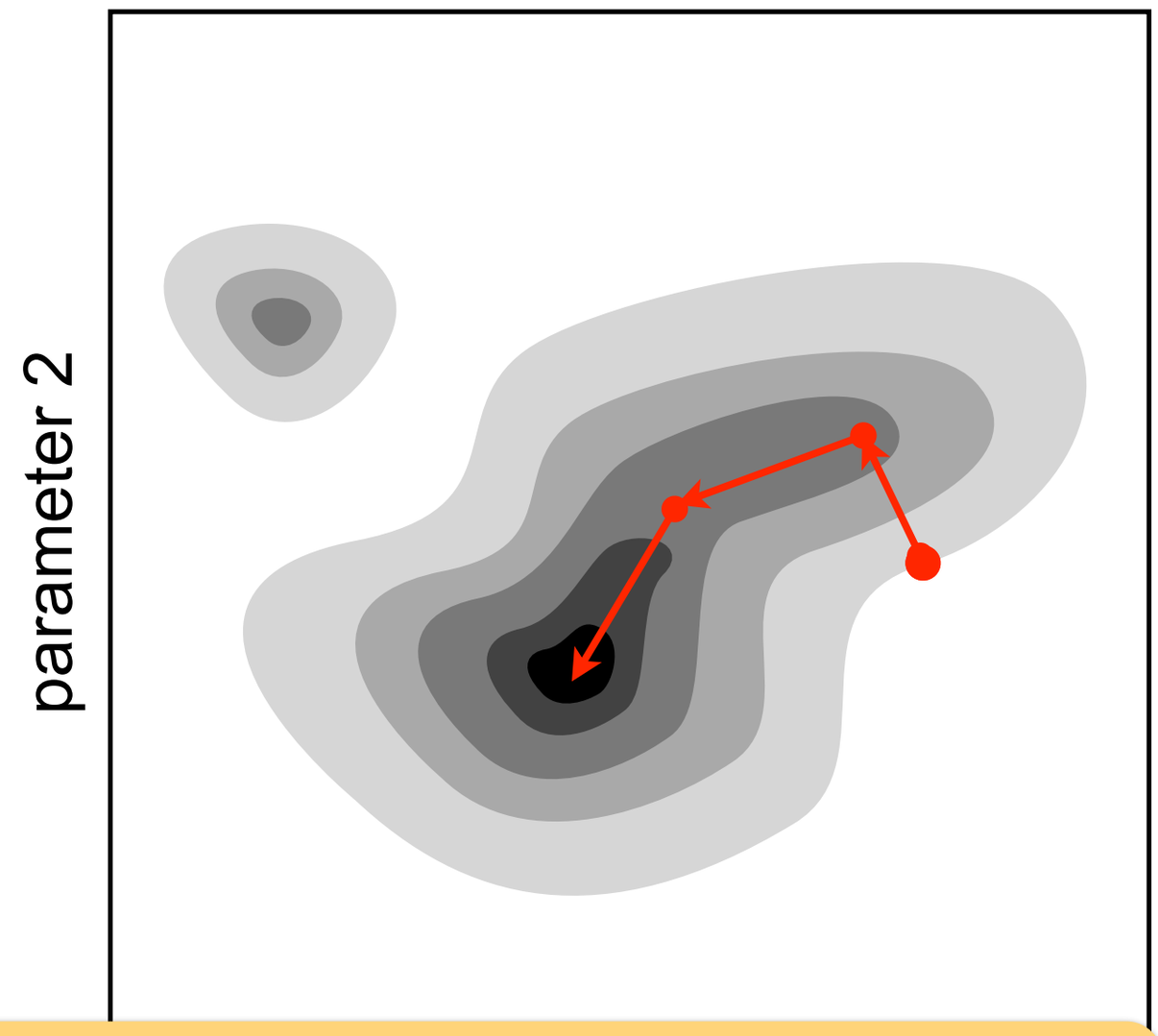
Sequential local optimisation

Goal: Construct a sequence of points along which the fit improves, meaning that the objective functions decreases

Model-data comparison



Objective function landscape



Evaluation of the objective function gradient?

How can sensitivities and gradients be computed?

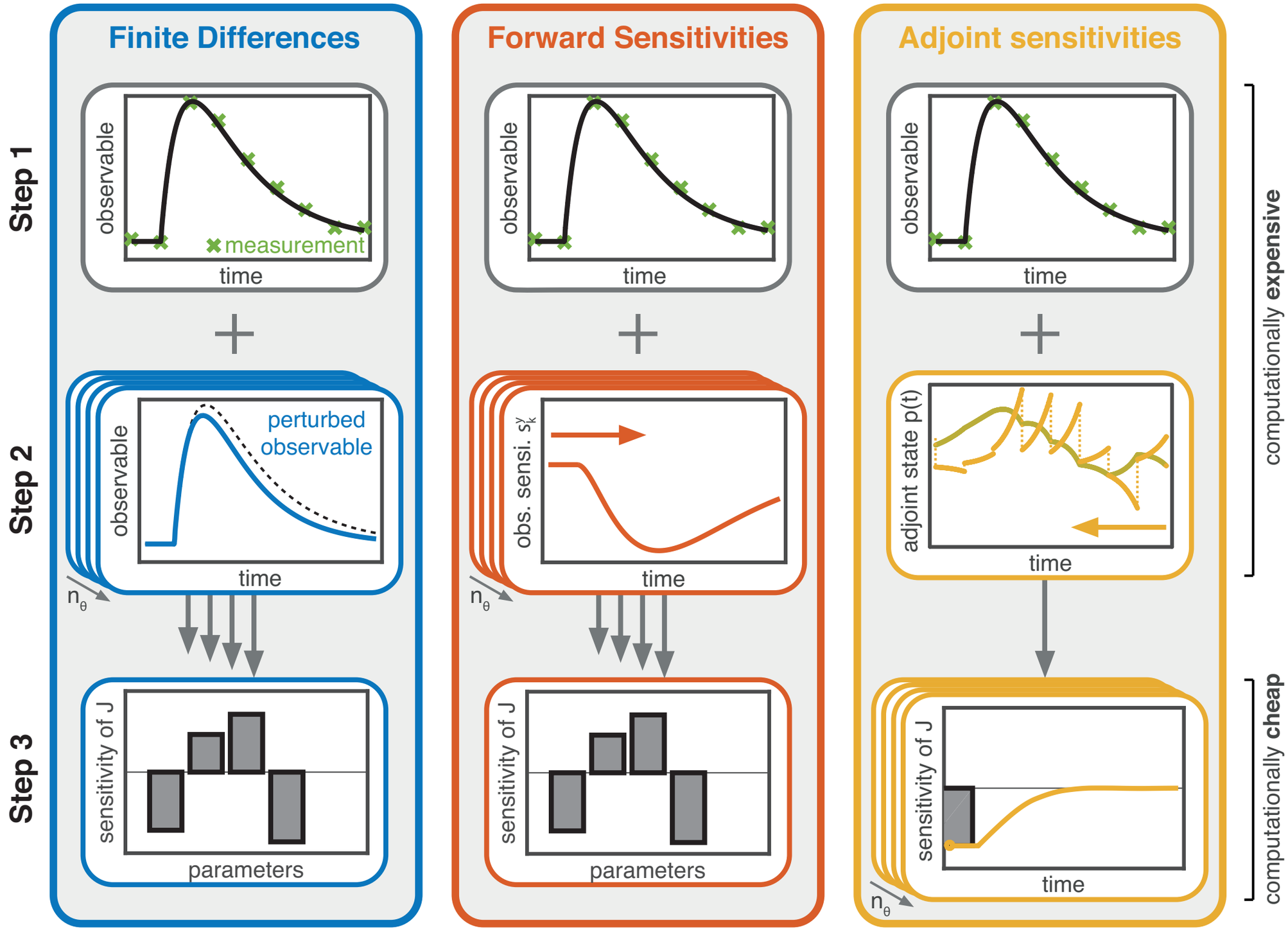
gradient

=

change of objective function

with respect to the parameters

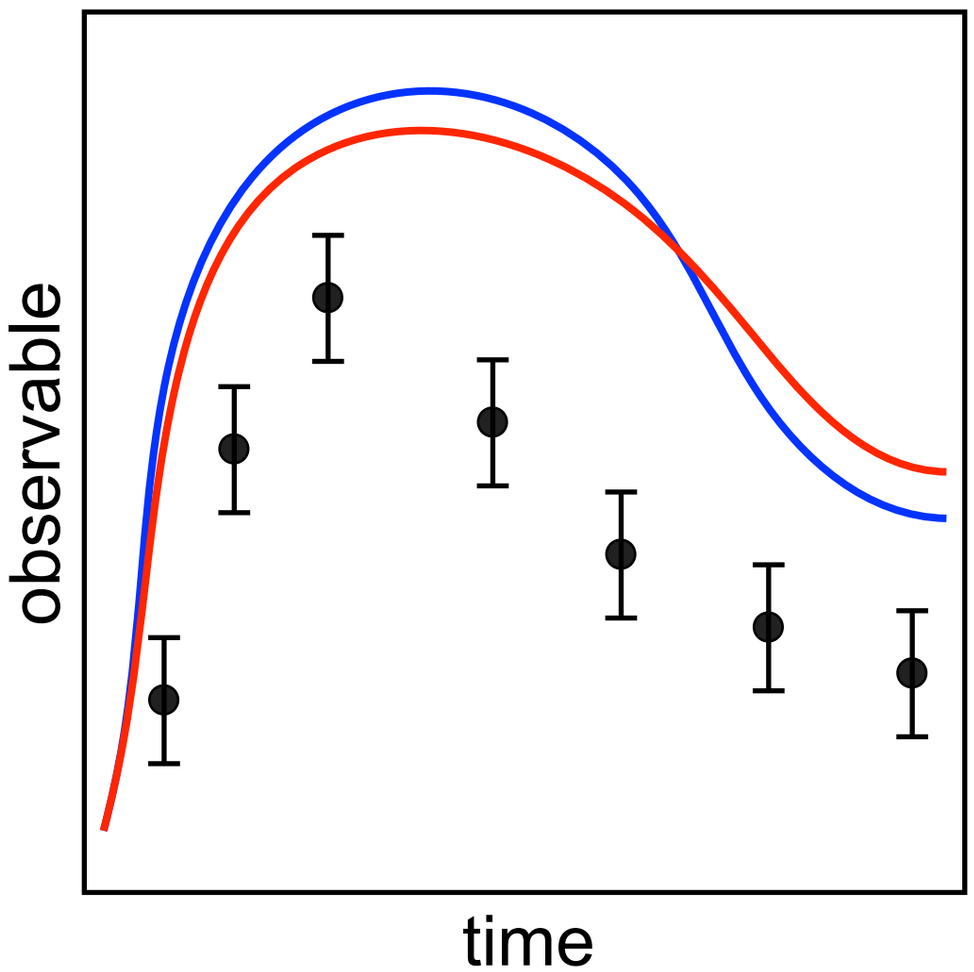
Gradient evaluation methods



Fröhlich el al. Scalable parameter estimation for genome-scale biochemical reaction networks. PLoS Computational Biology, 13(1):e1005331, 2017.

Finite differences

1. Calculation of objective function for parameter θ
2. Calculation of objective function for parameter $\theta + \Delta e_i$
3. Calculation of gradient as the normalised difference $g_i(\theta)$



$$g_i(\theta) = \frac{J(\theta + \Delta e_i) - J(\theta)}{\Delta}$$

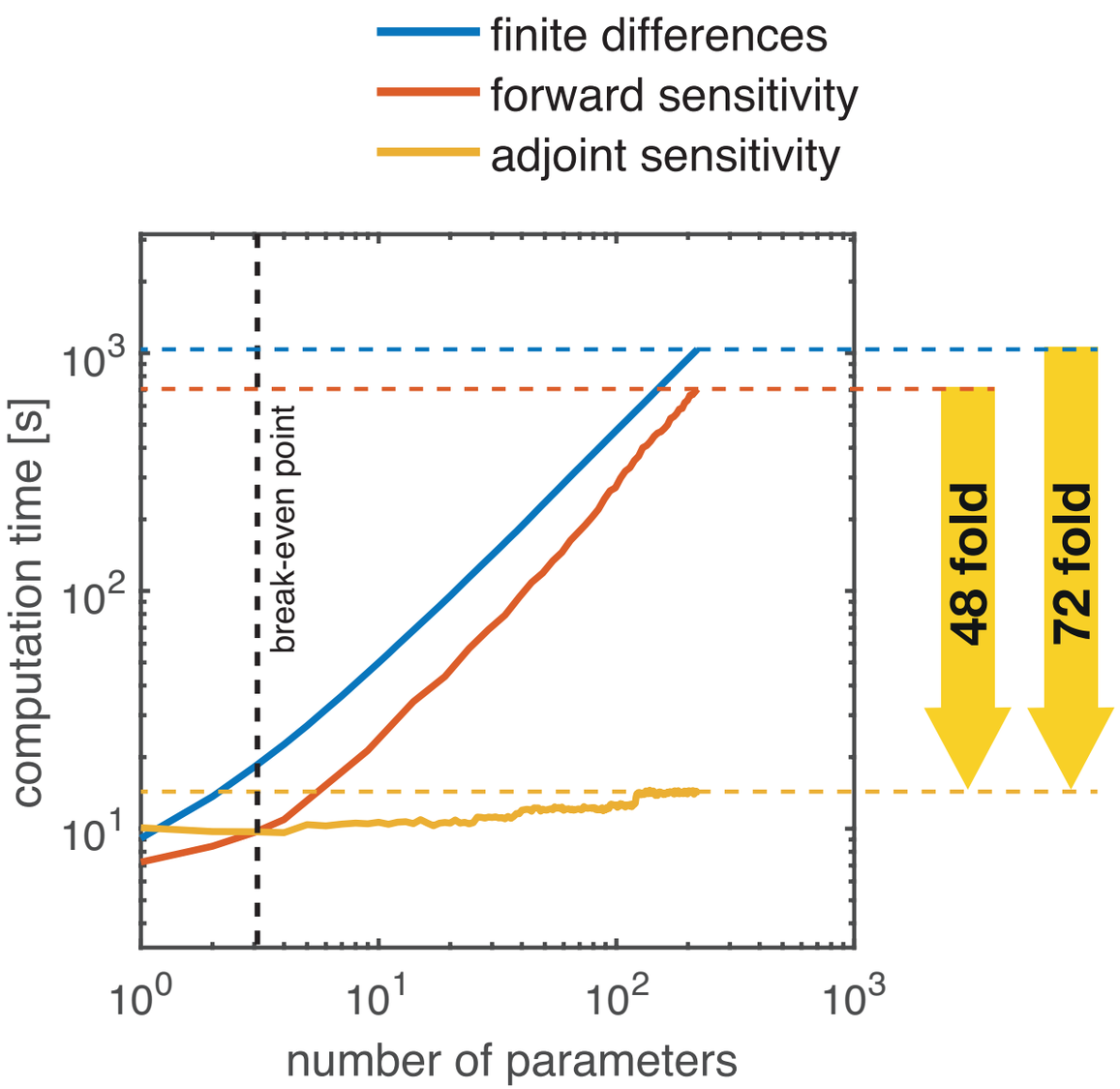
What to do?

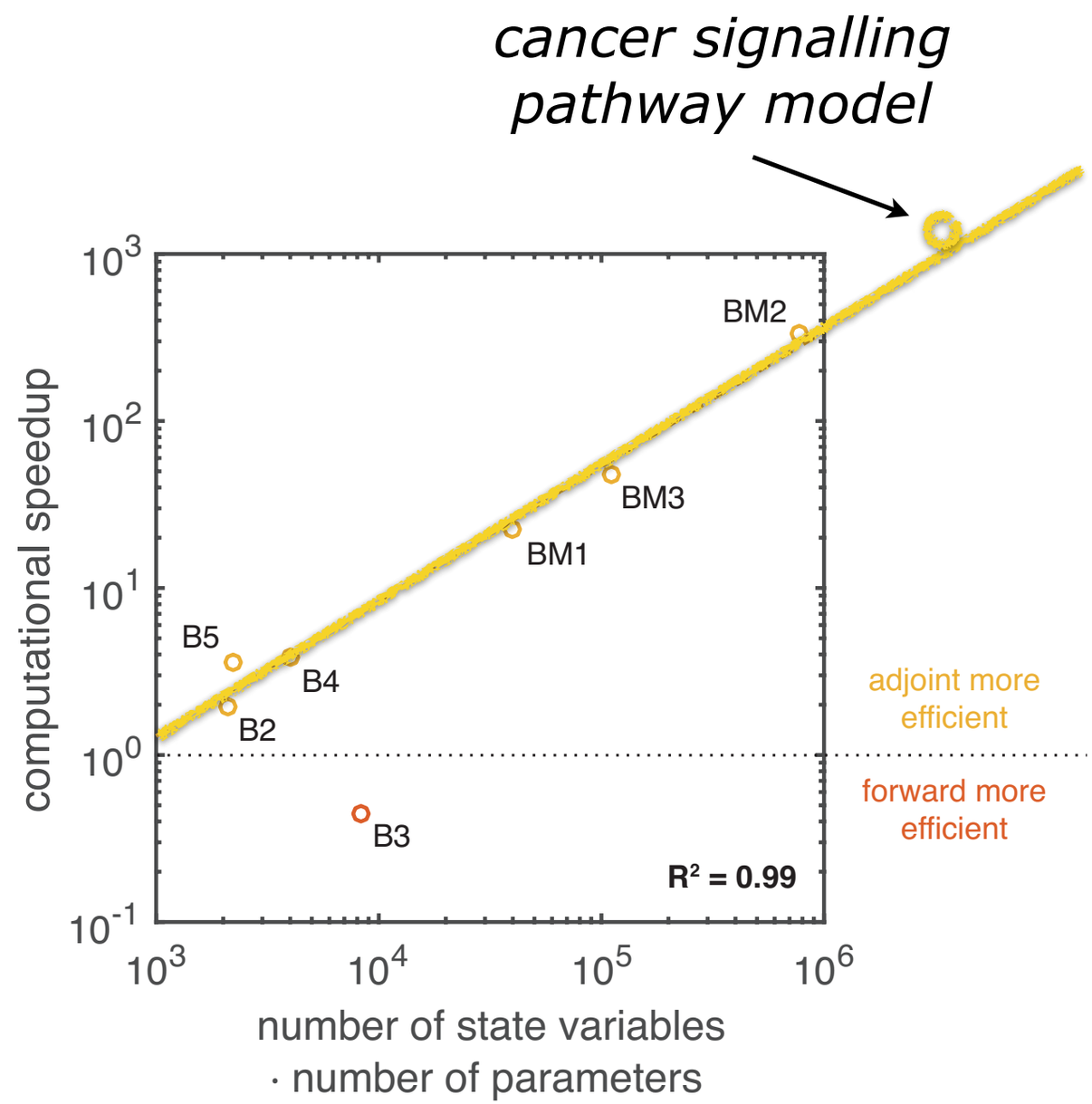
$g_i(\theta) > 0$: decrease parameter value

$g_i(\theta) < 0$: increase parameter value

$g_i(\theta) = 0$: stop

Scalability





Adjoint methods facilitate scalable gradient evaluation.

Literature

Numerical solvers

- A. C. Hindmarsh et al., SUNDIALS: Suite of Nonlinear and Differential/Algebraic Equation Solvers. ACM Transactions on Mathematical Software. 31(3):363-396, 2005

Sensitivities and gradients

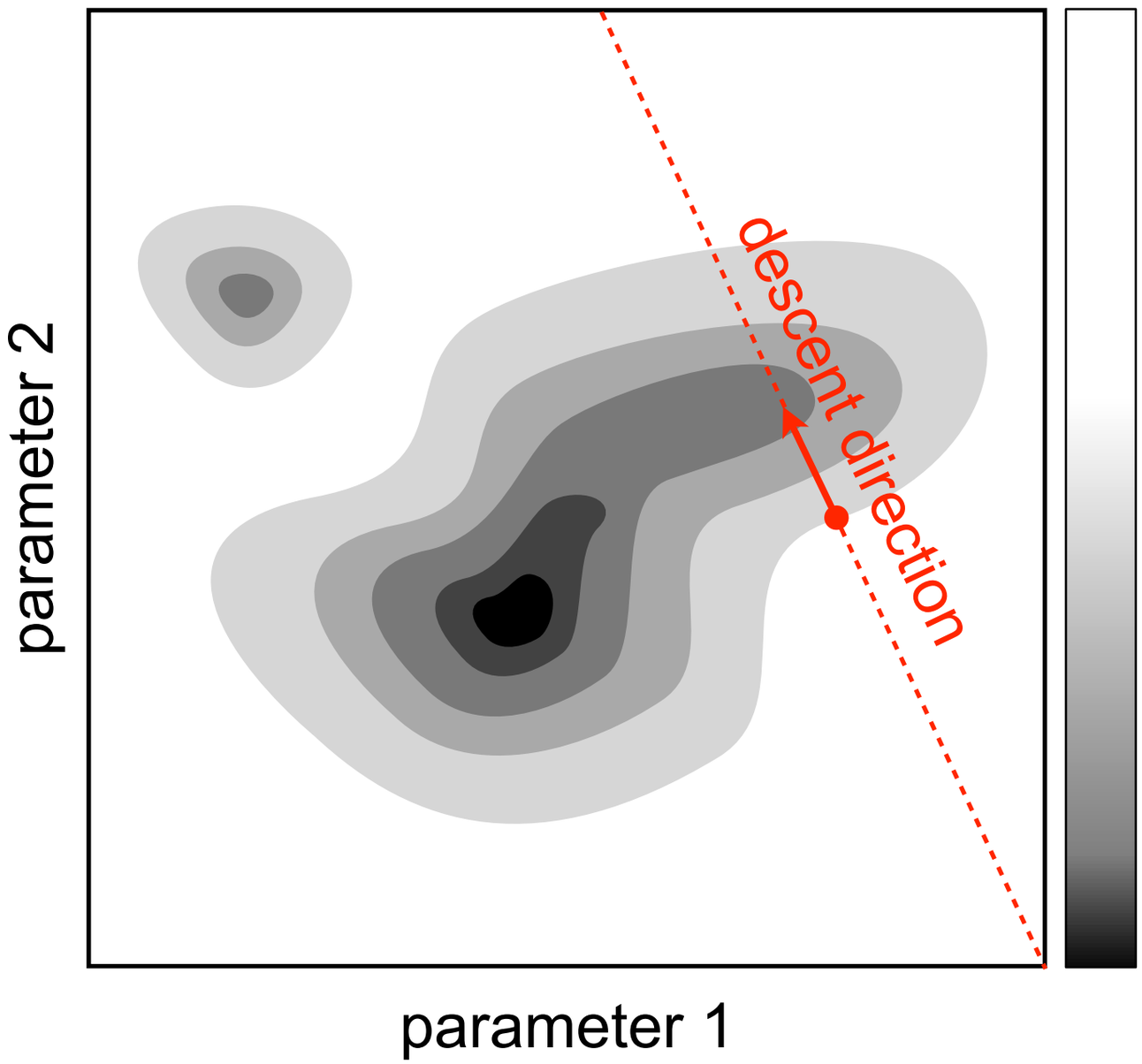
- A. Raue et al. Lessons learned from quantitative dynamical modeling in systems biology. PLoS ONE, 8(9):e74335, 2013.
- F. Fröhlich, B. Kaltenbacher, F. J. Theis, and J. Hasenauer. Scalable Parameter Estimation for Genome-Scale Biochemical Reaction Networks. PLoS Computational Biology, 13(1):e1005331, 2017.

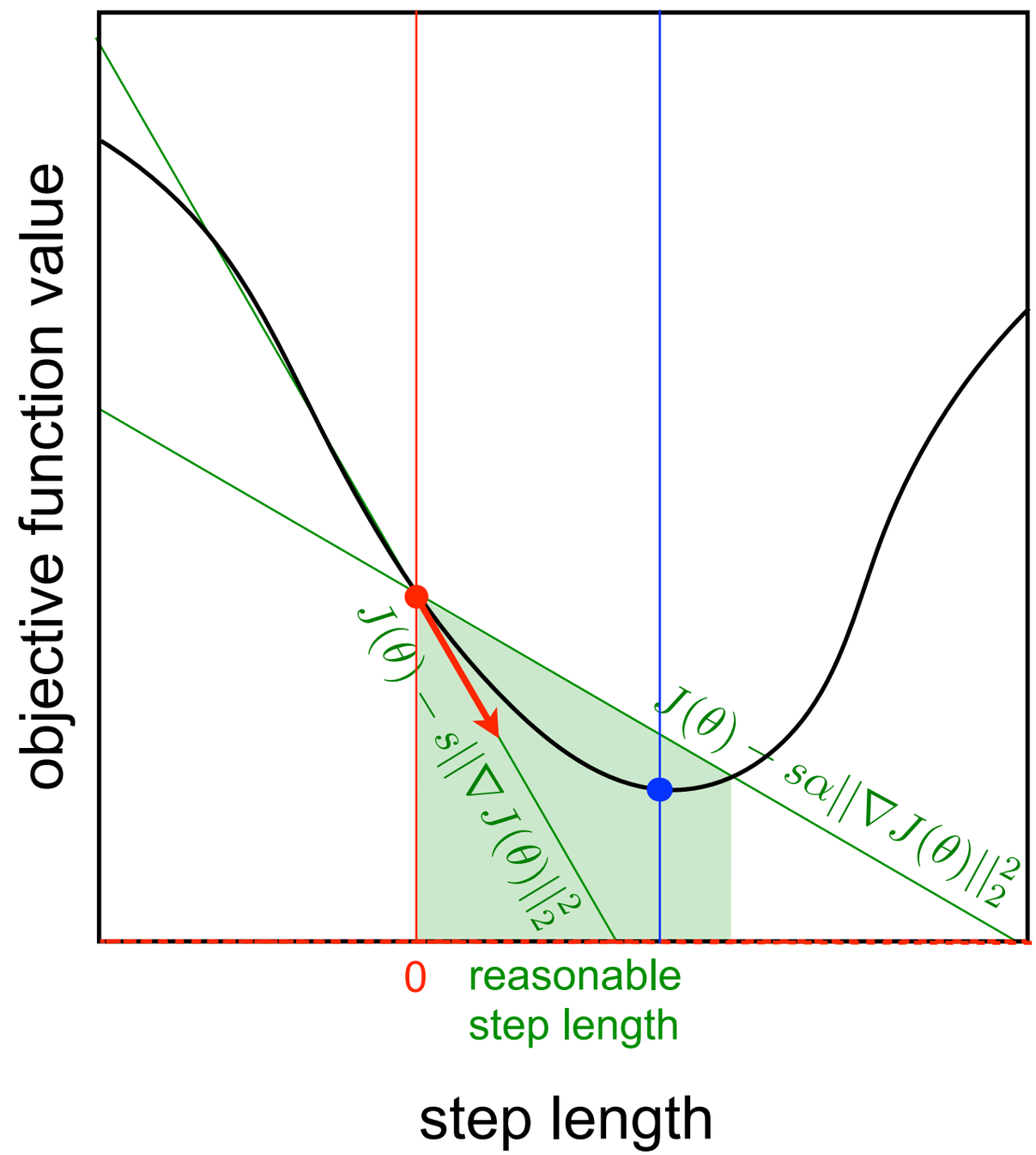
Evaluation of problem formulation

- H. Hass, C. Loos, E. Raimundez Alvarez, J. Timmer, J. Hasenauer, and C. Kreutz, Benchmark Problems for Dynamic Modeling of Intracellular Processes. Bioinformatics, 2019.

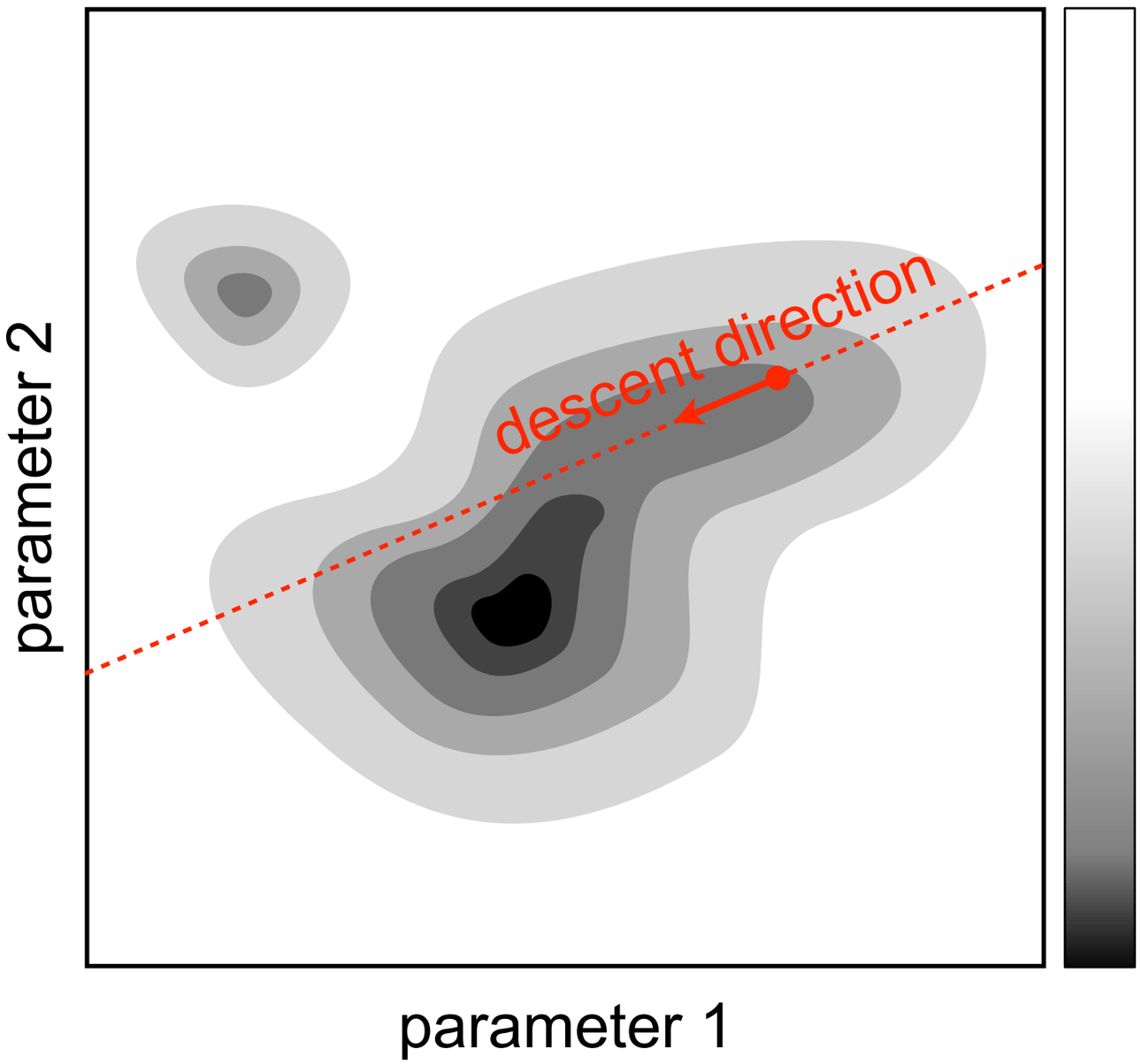
How does optimisation work?

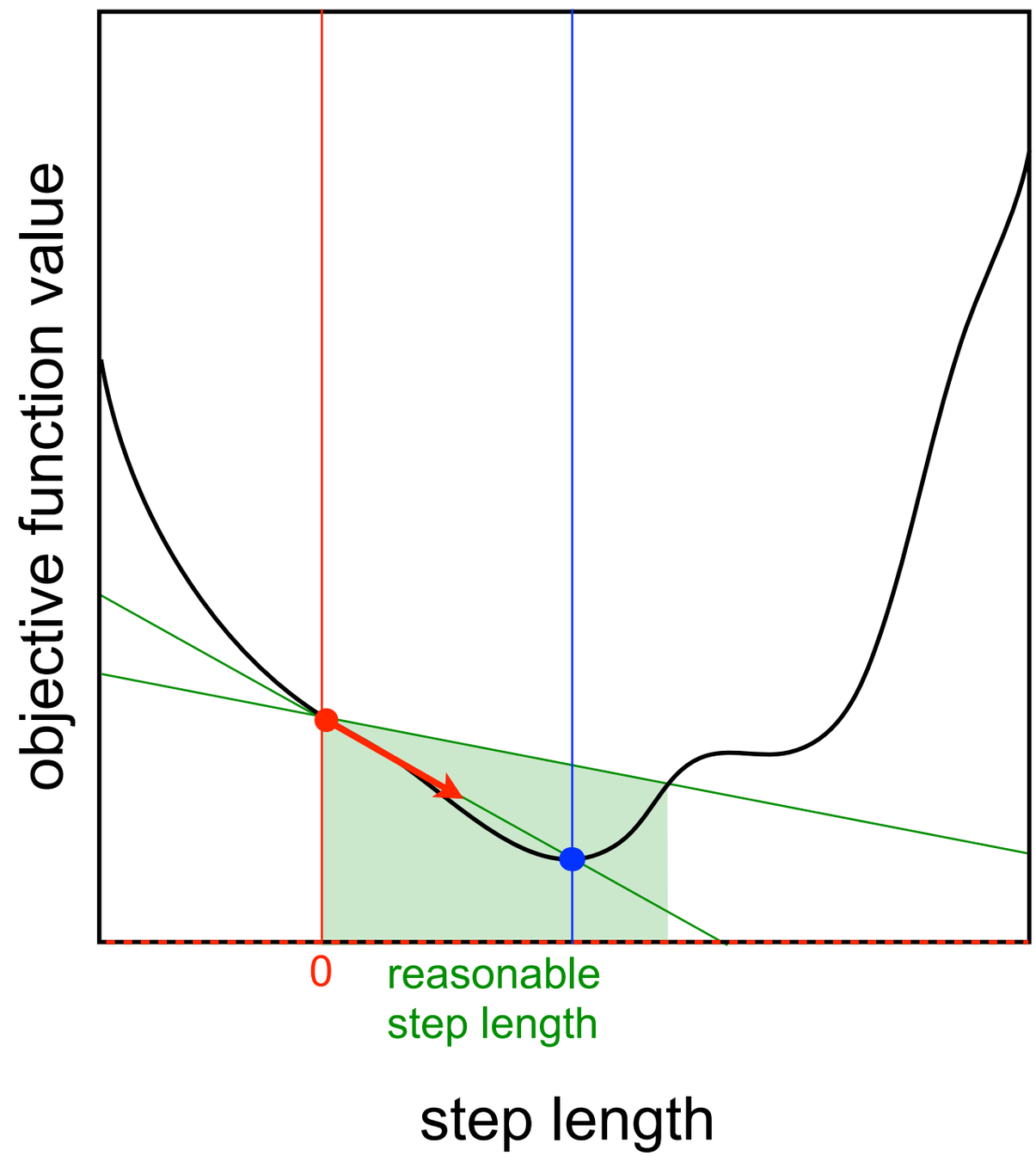
Line search method





Line search method





Selection of descent direction

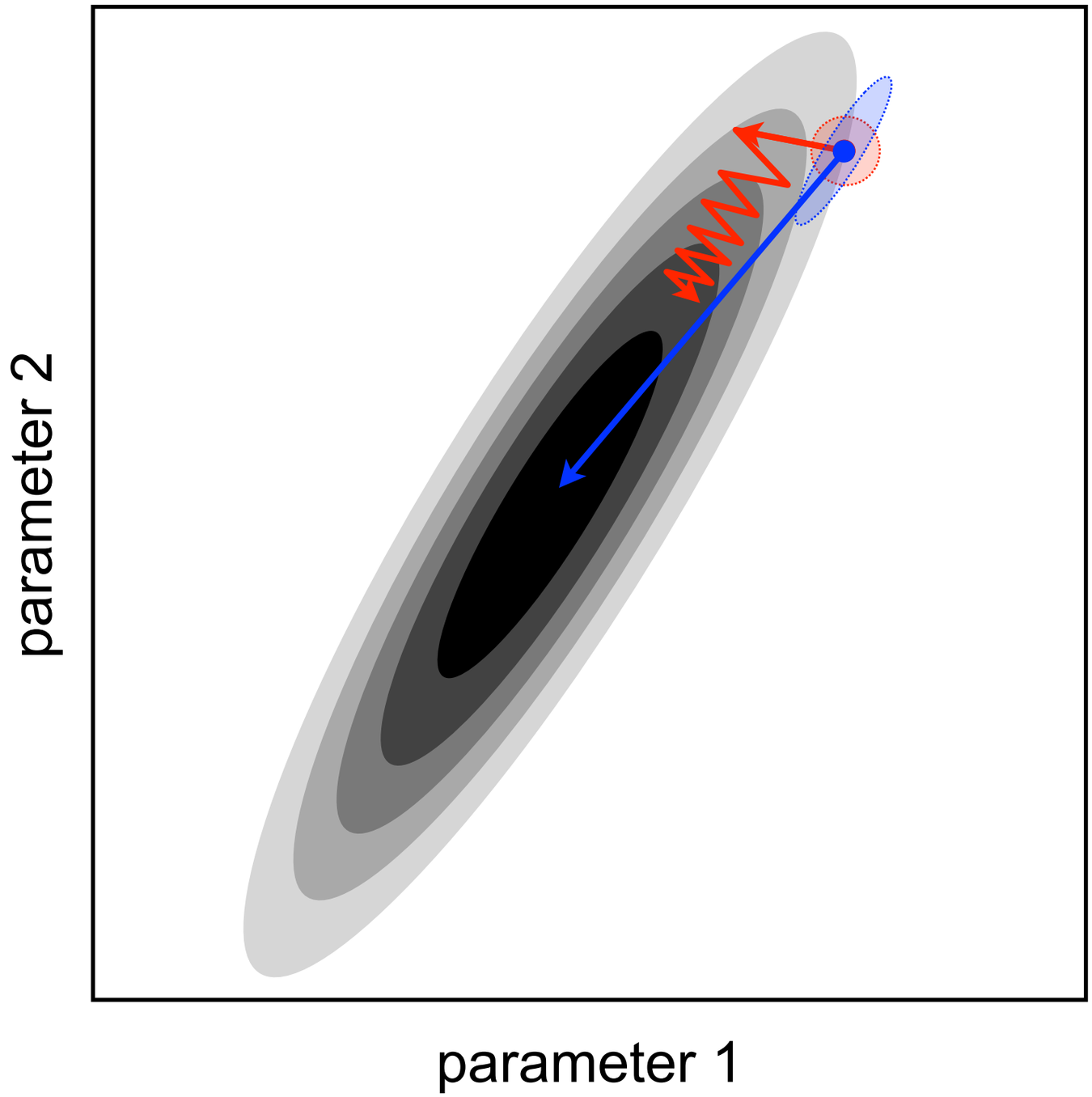
Gradient descent

$$d(\theta) = -g(\theta)$$

Steepest descent using local curvature

$$d(\theta) = -H(\theta)^{-1} g(\theta)$$

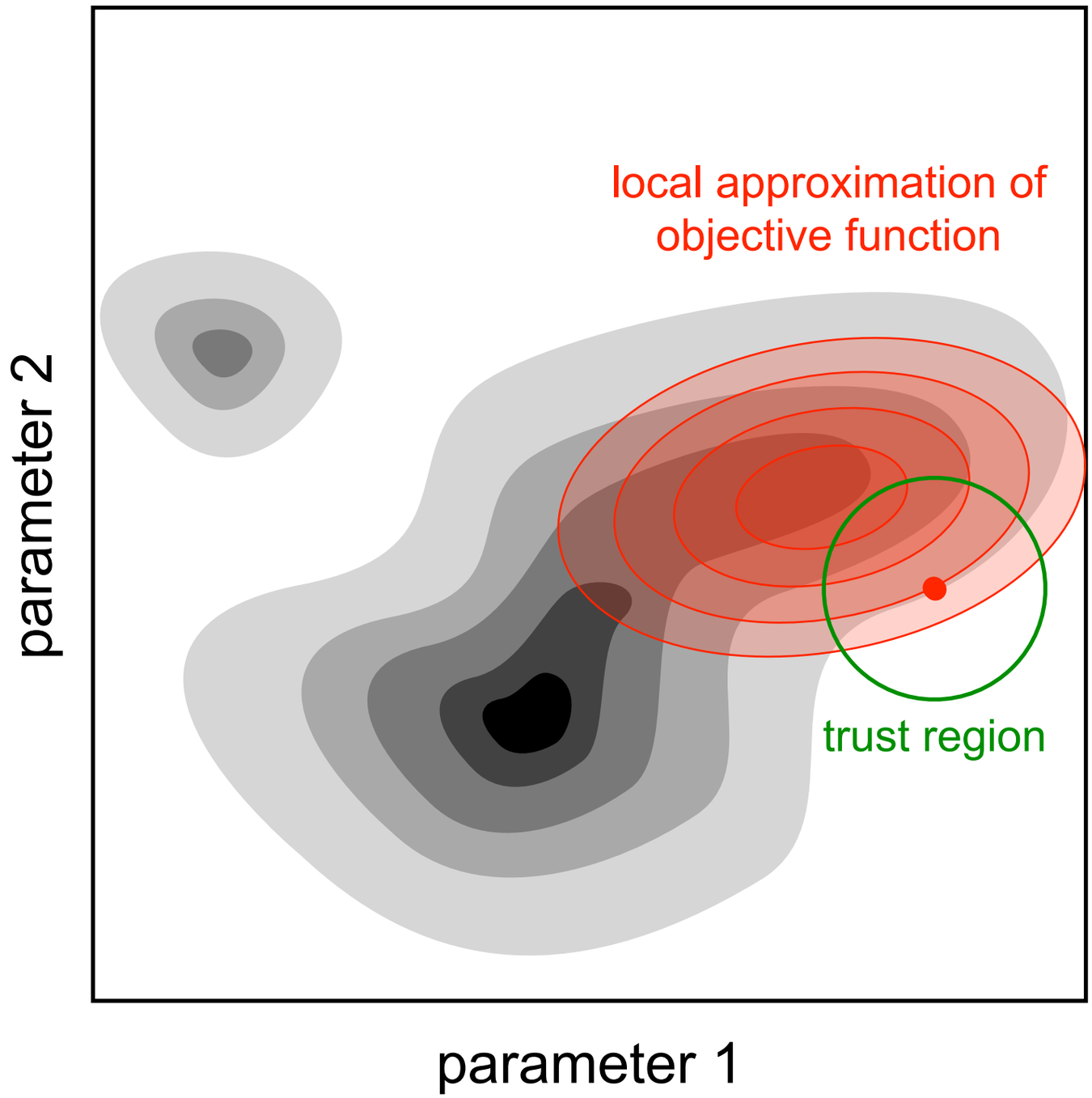
Remark: Steepest descent using local curvature converges faster but requires a positive-definite Hessian.



Trust-region method

Local approximation using Taylor series

$$\tilde{J}(\theta') = J(\theta) + g(\theta)^T(\theta' - \theta) + \frac{1}{2}(\theta' - \theta)^T H(\theta)(\theta' - \theta)$$



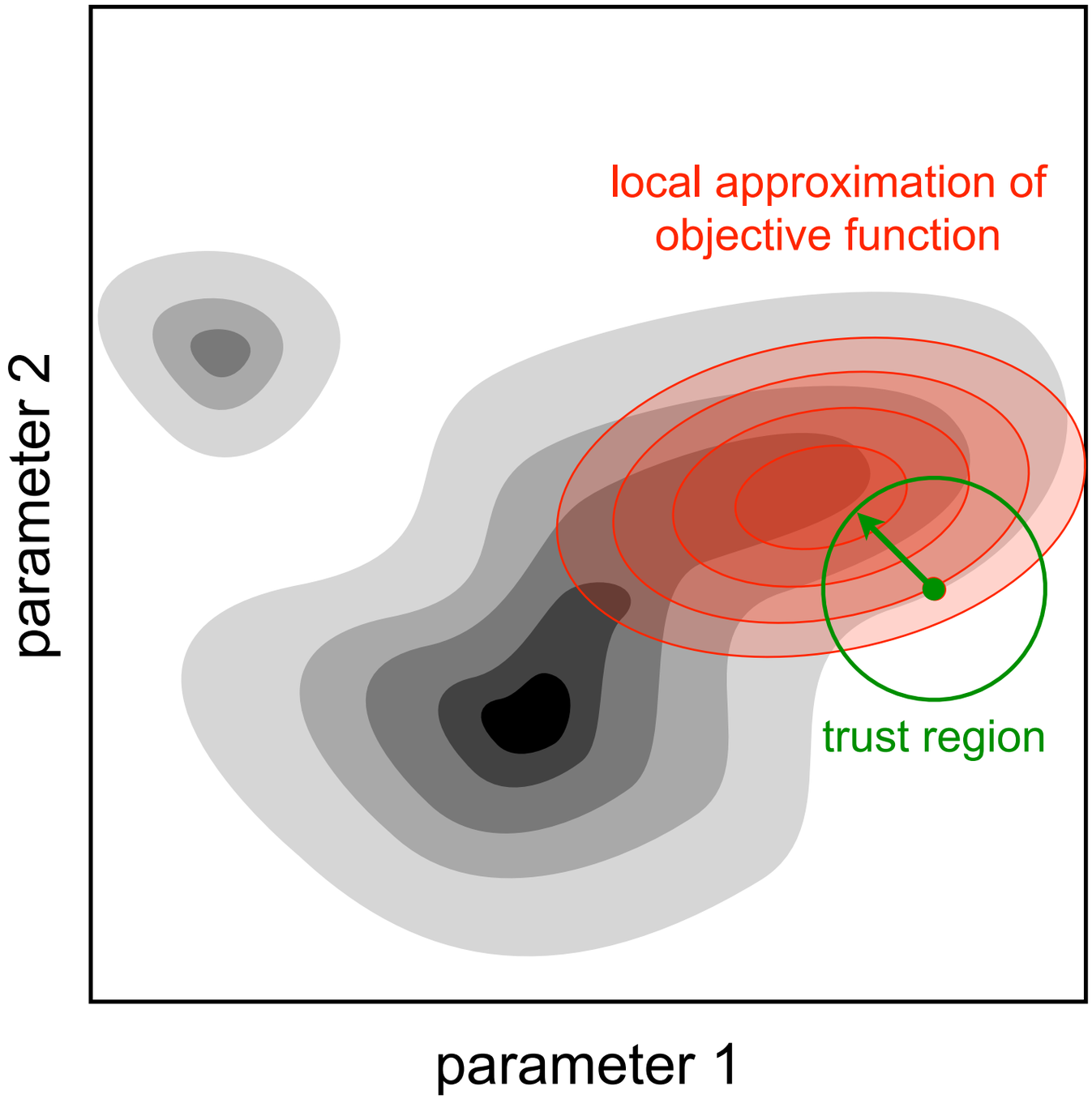
Trust-region method

Local approximation using Taylor series

$$\tilde{J}(\theta') = J(\theta) + g(\theta)^T(\theta' - \theta) + \frac{1}{2}(\theta' - \theta)^T H(\theta)(\theta' - \theta)$$

Trust-region subproblem

$$\min_{\theta' \in \mathrm{TR}} \tilde{J}(\theta')$$



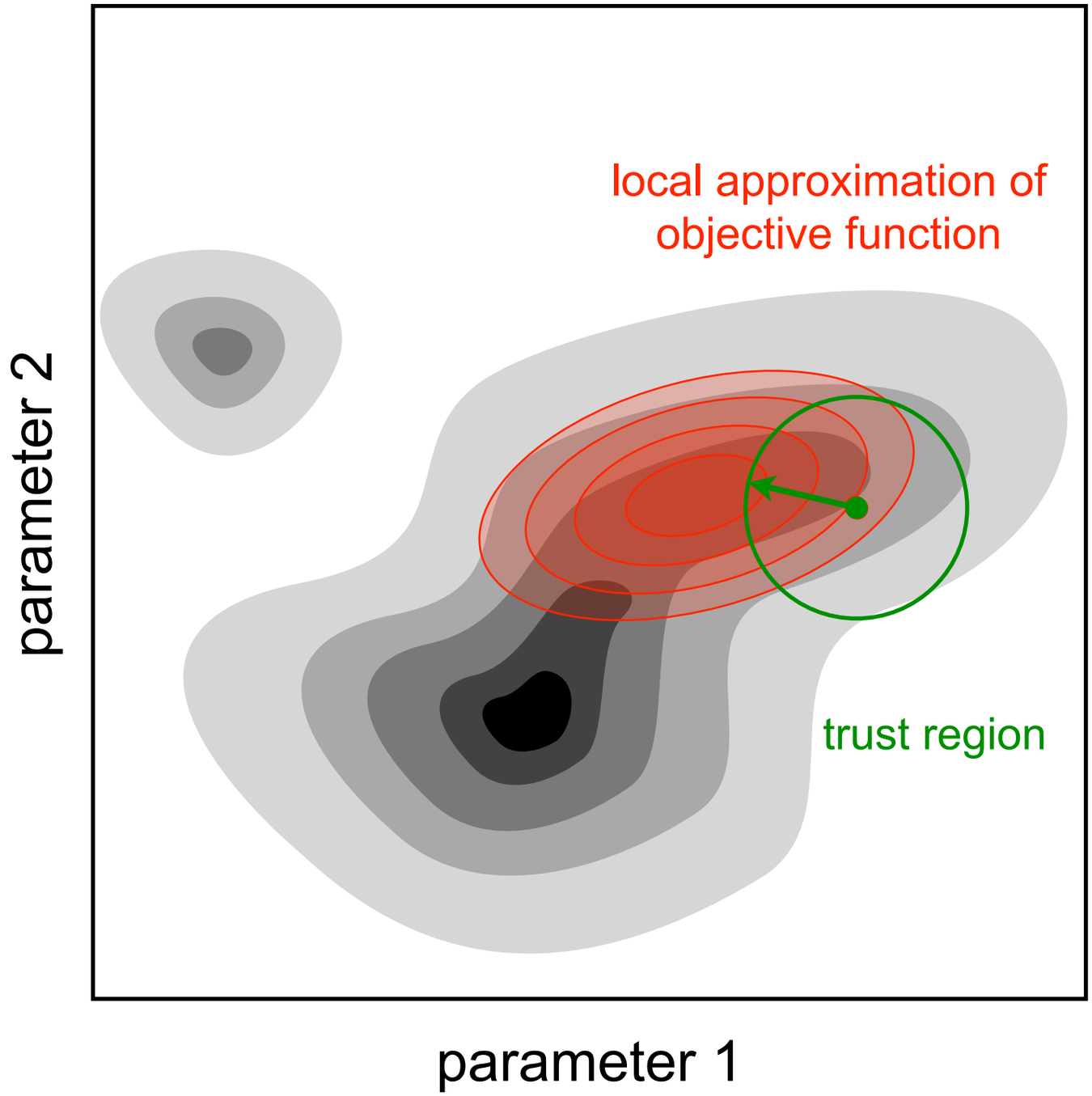
Trust-region method

Local approximation using Taylor series

$$\tilde{J}(\theta') = J(\theta) + g(\theta)^T(\theta' - \theta) + \frac{1}{2}(\theta' - \theta)^T H(\theta)(\theta' - \theta)$$

Trust-region subproblem

$$\min_{\theta' \in \mathrm{TR}} \tilde{J}(\theta')$$



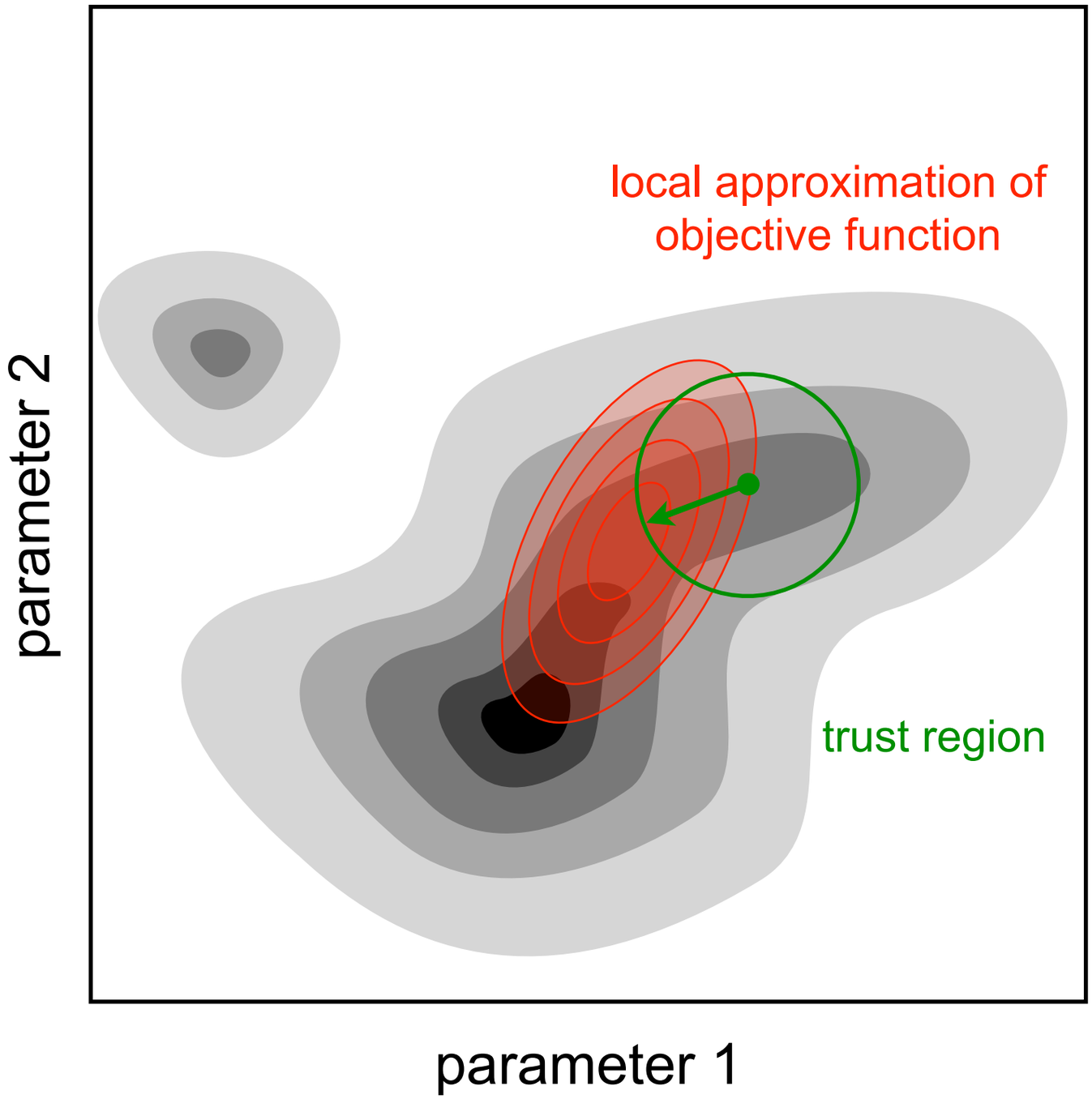
Trust-region method

Local approximation using Taylor series

$$\tilde{J}(\theta') = J(\theta) + g(\theta)^T(\theta' - \theta) + \frac{1}{2}(\theta' - \theta)^T H(\theta)(\theta' - \theta)$$

Trust-region subproblem

$$\min_{\theta' \in \mathrm{TR}} \tilde{J}(\theta')$$



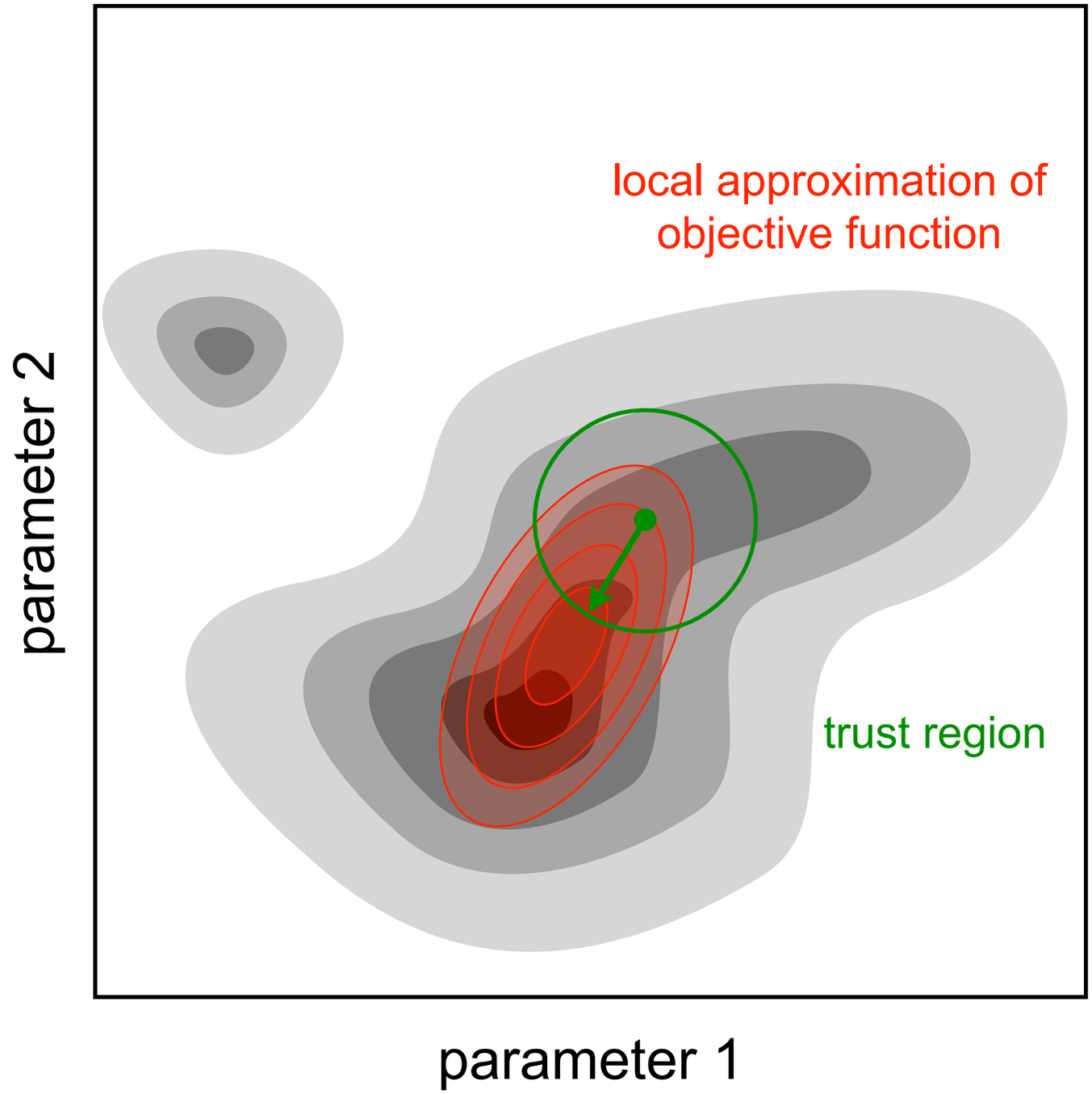
Trust-region method

Local approximation using Taylor series

$$\tilde{J}(\theta') = J(\theta) + g(\theta)^T(\theta' - \theta) + \frac{1}{2}(\theta' - \theta)^T H(\theta)(\theta' - \theta)$$

Trust-region subproblem

$$\min_{\theta' \in \mathrm{TR}} \tilde{J}(\theta')$$



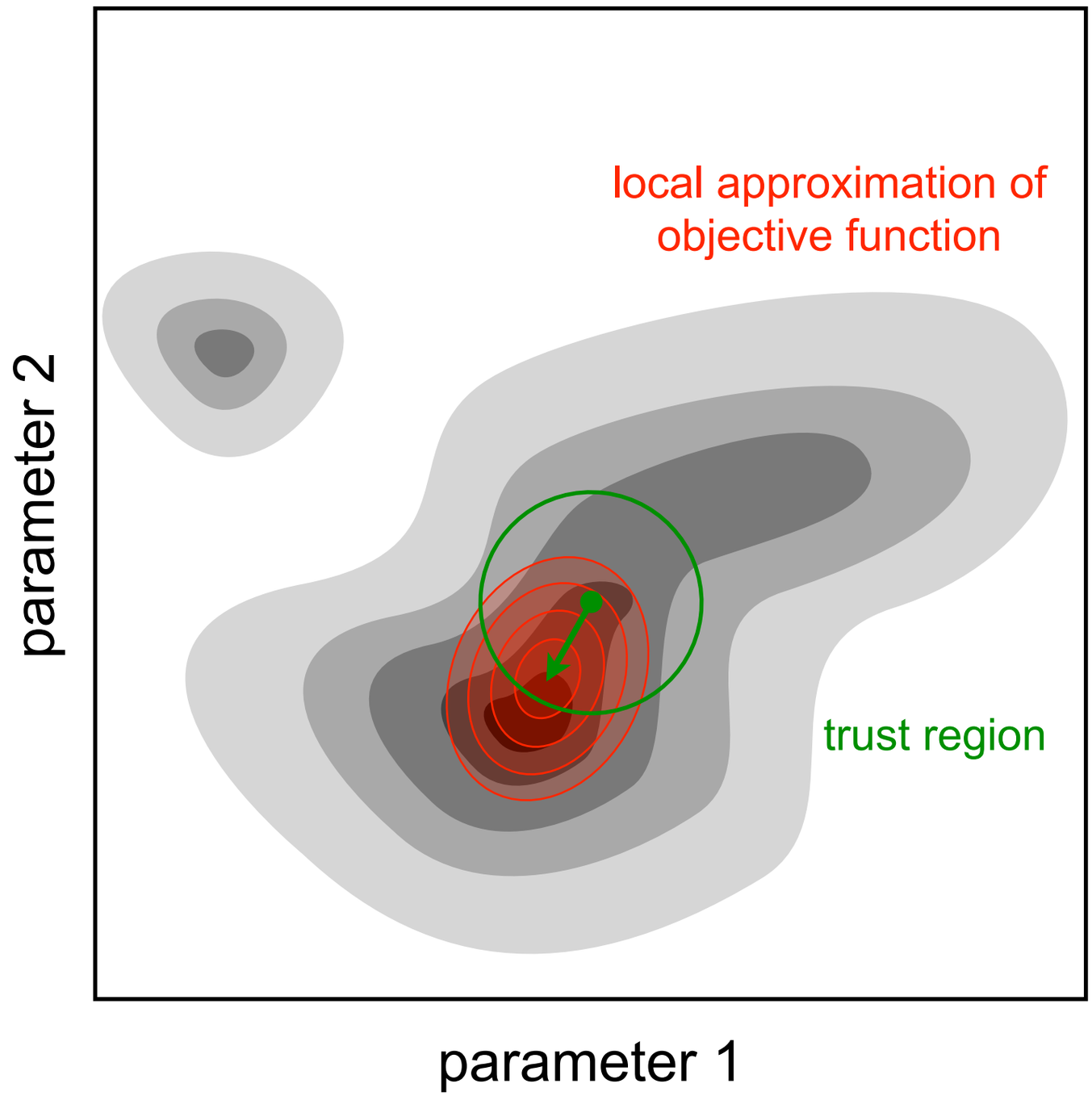
Trust-region method

Local approximation using Taylor series

$$\tilde{J}(\theta') = J(\theta) + g(\theta)^T(\theta' - \theta) + \frac{1}{2}(\theta' - \theta)^T H(\theta)(\theta' - \theta)$$

Trust-region subproblem

$$\min_{\theta' \in \mathrm{TR}} \tilde{J}(\theta')$$



Trust-region method

Local approximation using Taylor series

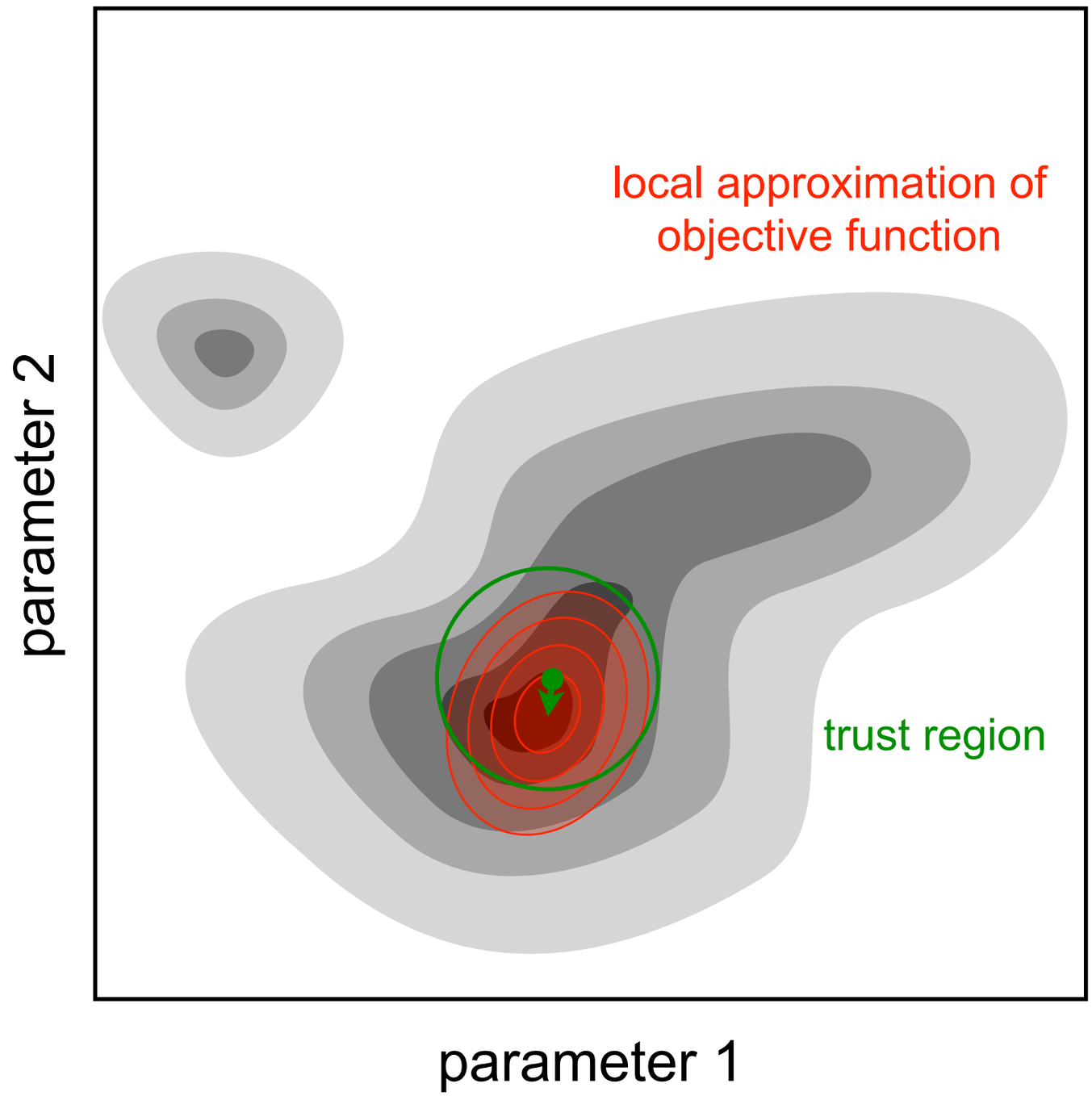
$$\tilde{J}(\theta') = J(\theta) + g(\theta)^T(\theta' - \theta) + \frac{1}{2}(\theta' - \theta)^T H(\theta)(\theta' - \theta)$$

Trust-region subproblem

$$\min_{\theta' \in \mathrm{TR}} \tilde{J}(\theta')$$

Remark:

- Approach allows conceptually for negative and indefinite Hessians.
- The solution of the subproblem is usually approximated.

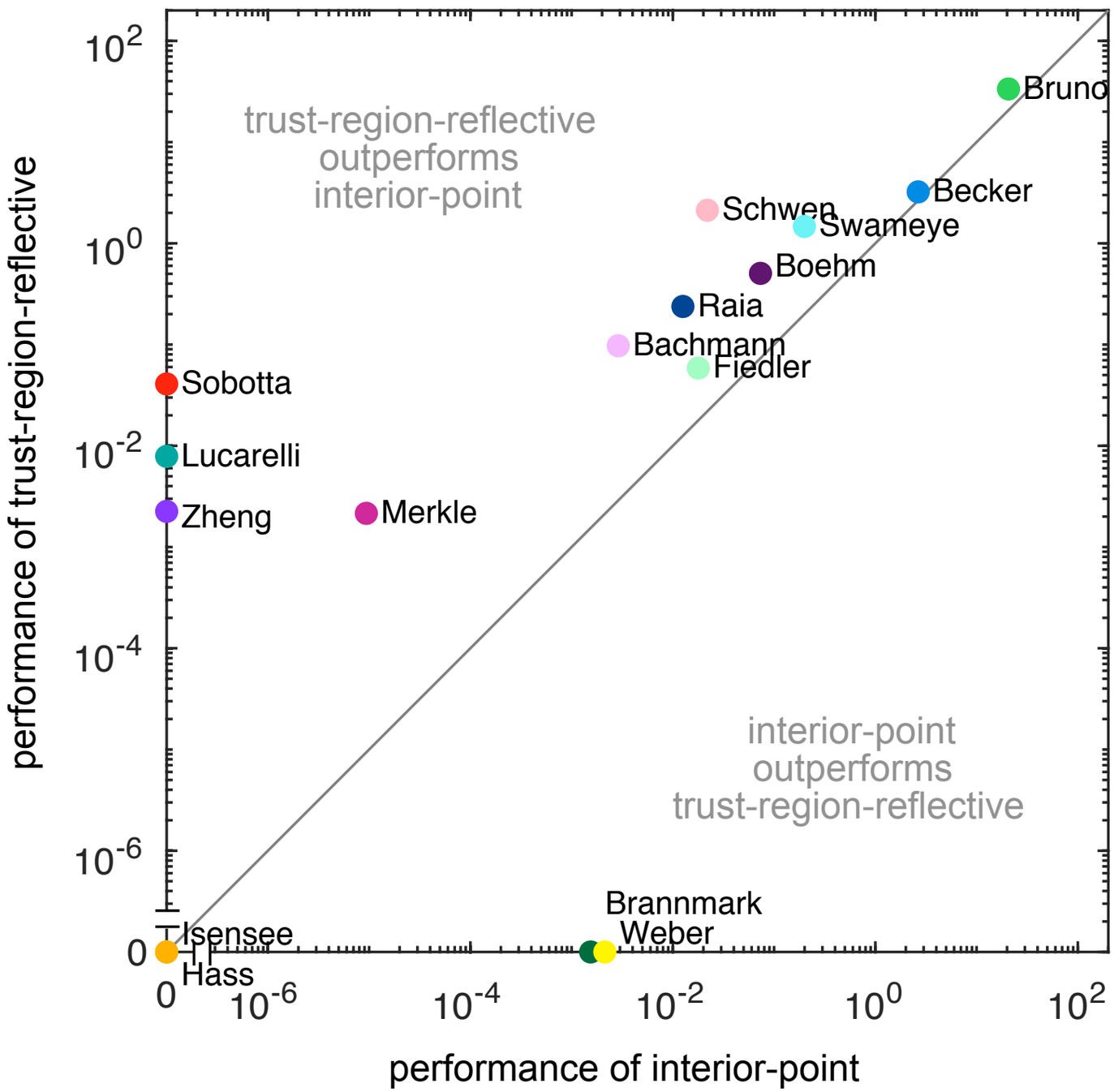


No free lunch theorem

"…what an algorithm gains in performance on one class of problems is necessarily offset by its performance on the remaining problems."

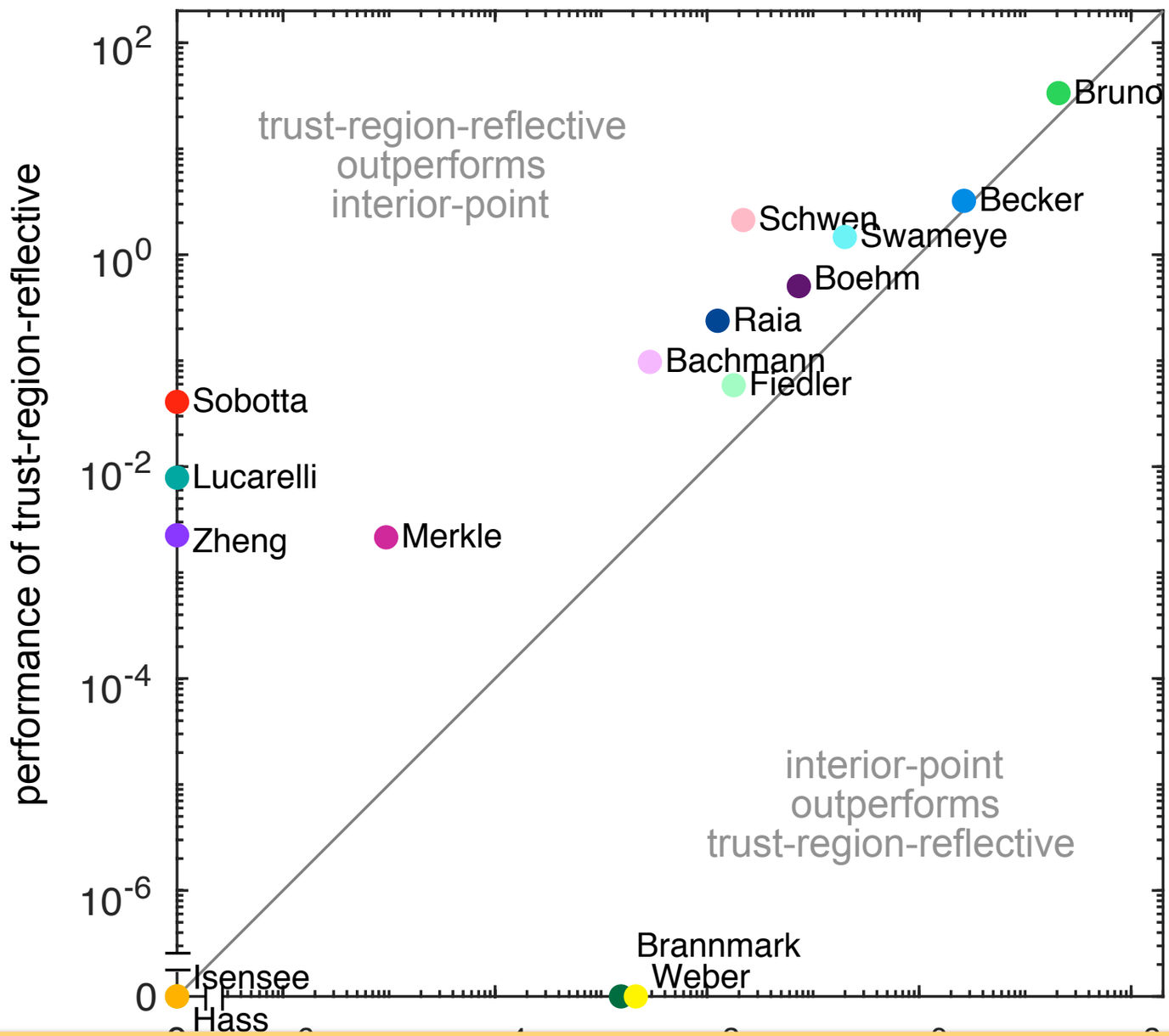
D. H. Wolpert and W. G. Macready. No free lunch theorems for optimization. IEEE Trans. Evol. Comput., 1(1):67-82,1997.

No free lunch theorem



Hass et al. Benchmark Problems for Dynamic Modeling of Intracellular Processes. *Bioinformatics*, 2019.

No free lunch theorem



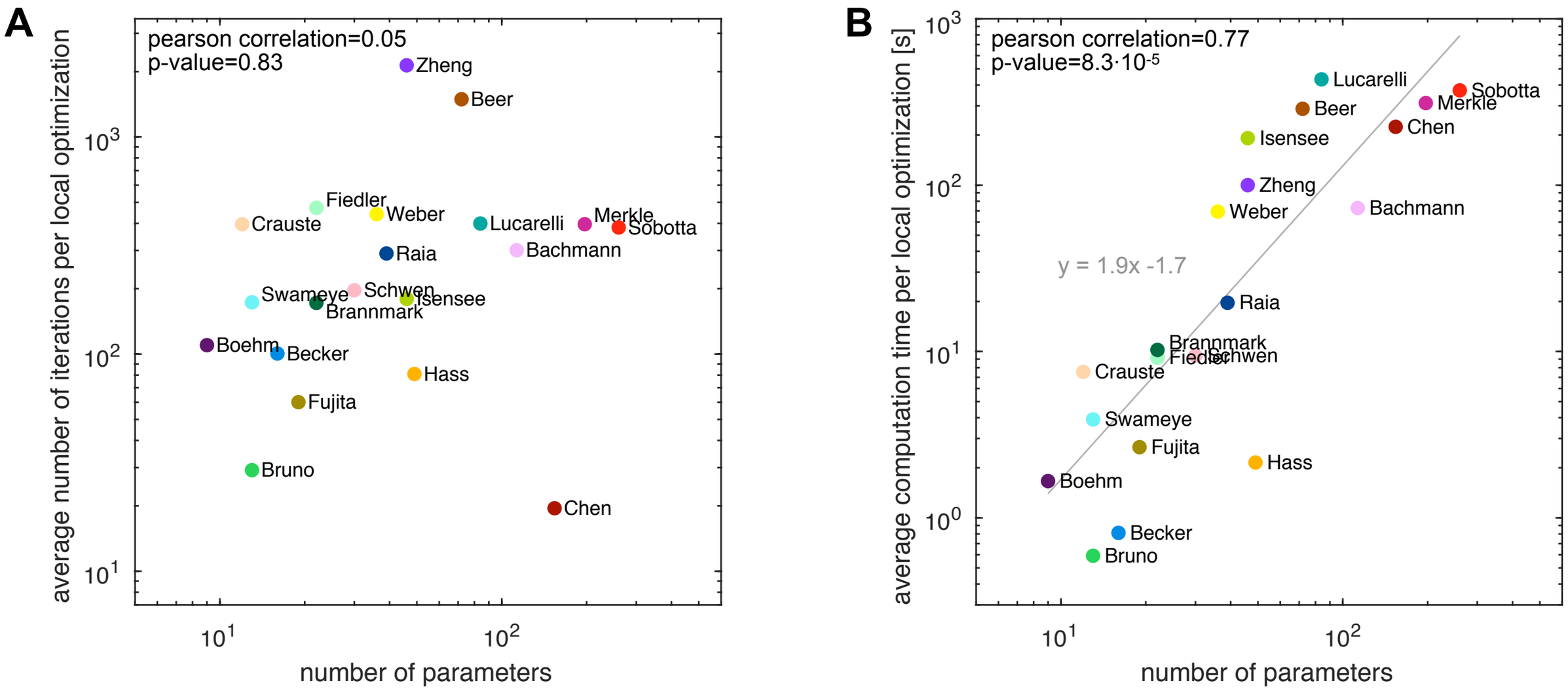
There is a free lunch for continues problems and the choice of the proper optimiser is crucial!

Number of iterations for local optimisers

"… the number of Newton steps hardly grows […] with m [the number of constraints - *author's note*] (or any other parameter, in fact)."

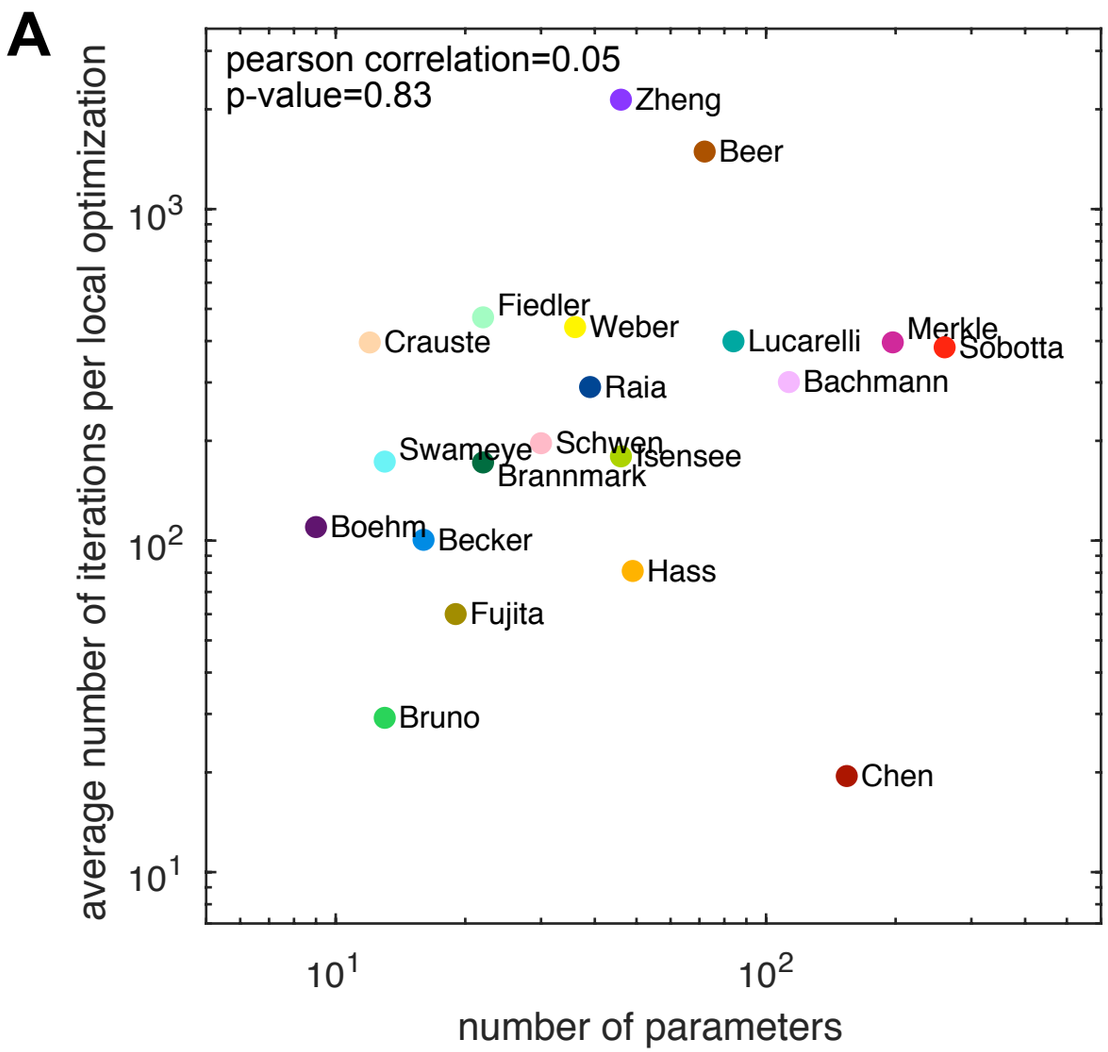
S. Boyd and L. Vandenberghe. Convex Optimisation, Cambridge University Press, UK, 2004.

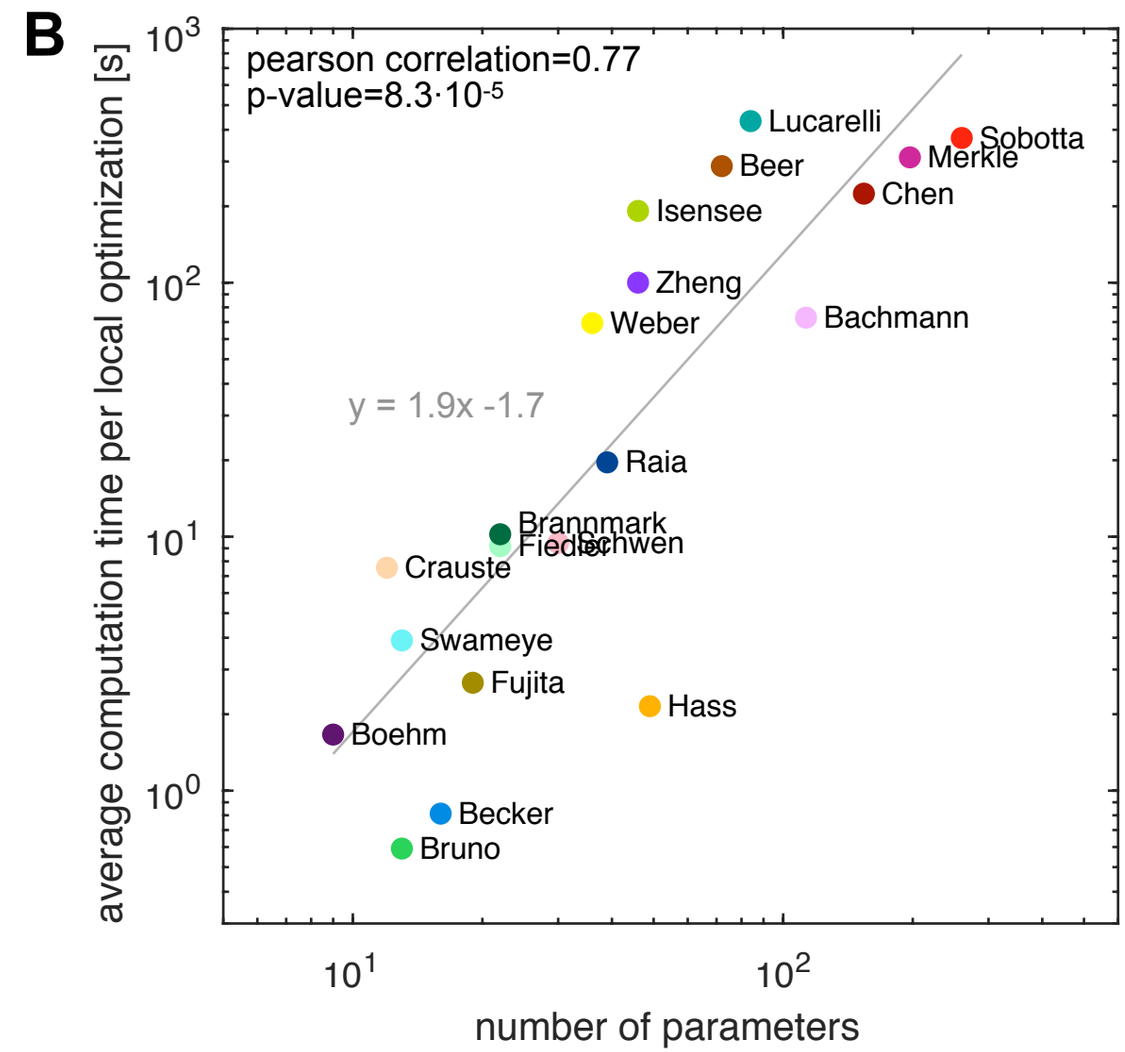
No free lunch theorem



Hass et al. Benchmark Problems for Dynamic Modeling of Intracellular Processes. *Bioinformatics*, 2019.

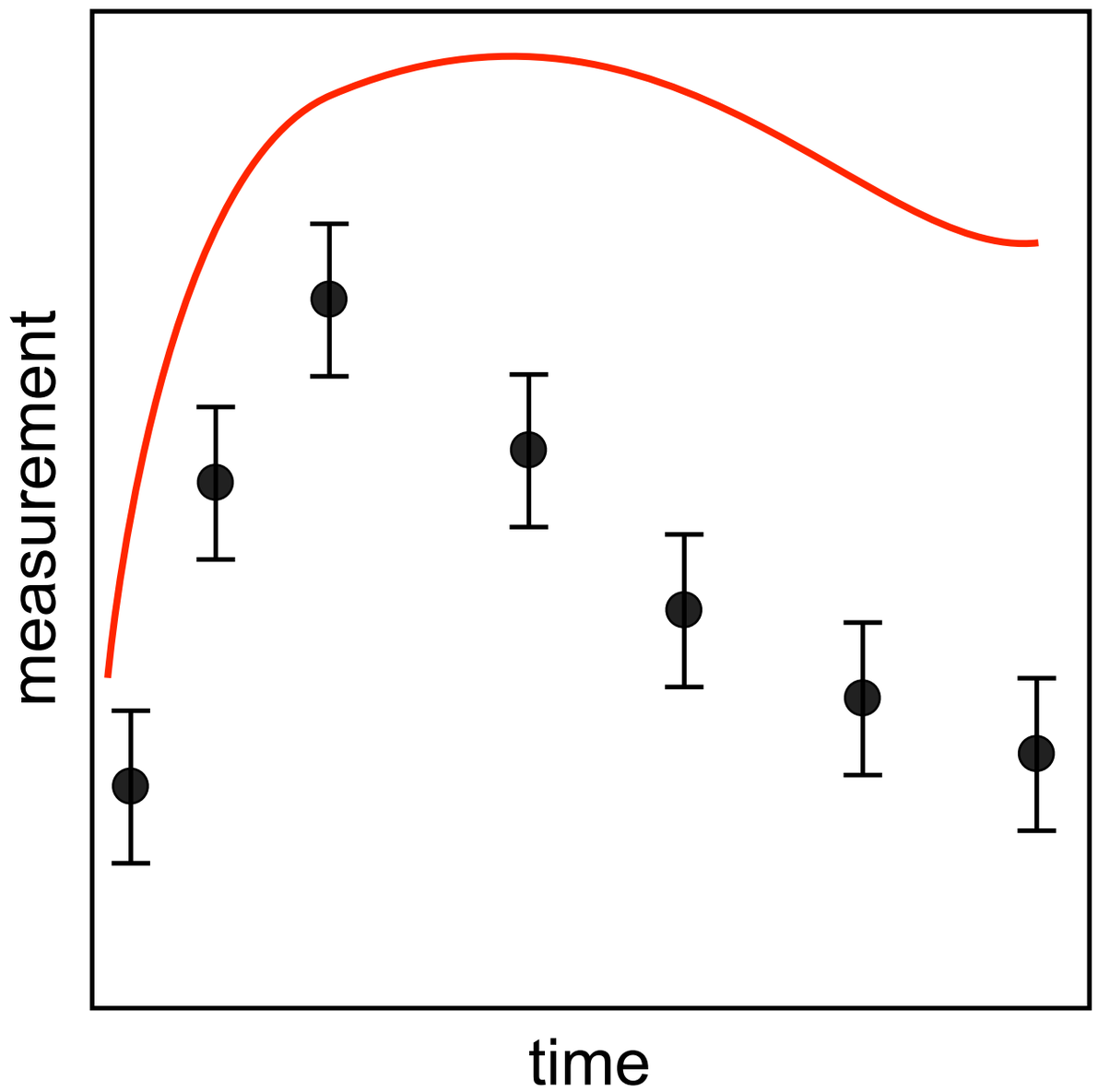
No free lunch theorem

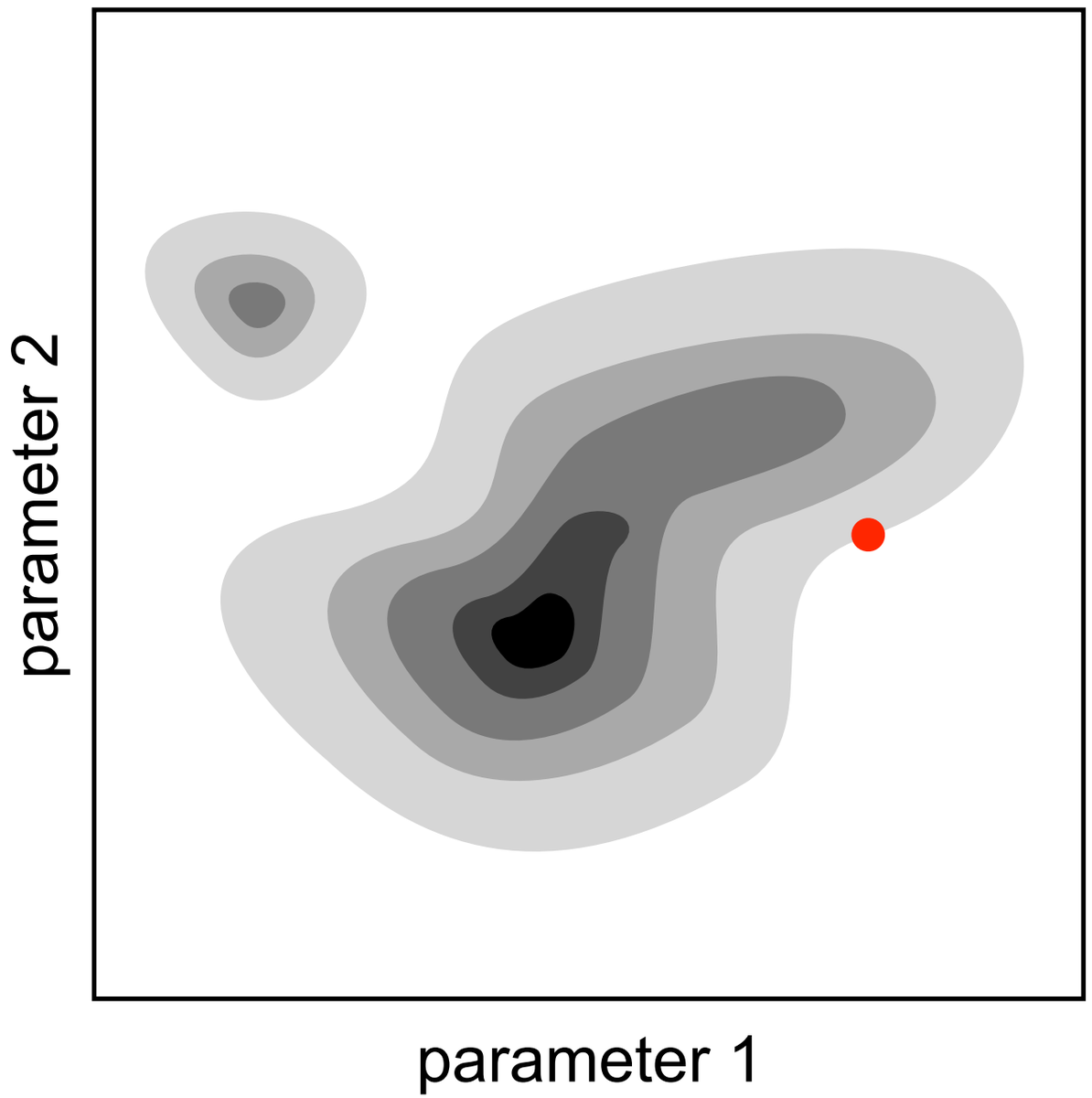




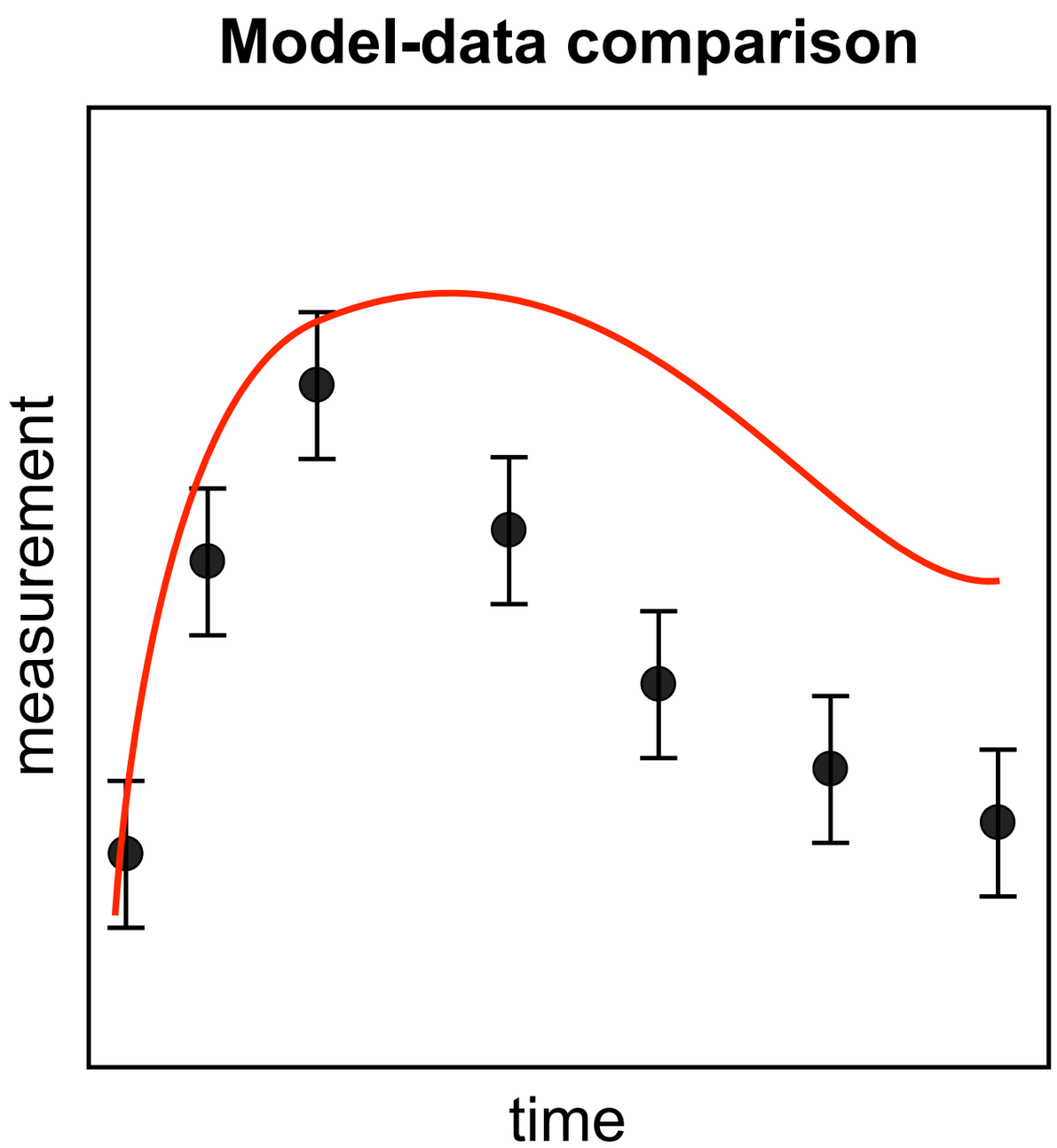
Favourable scaling is crucial for large-scale problems.

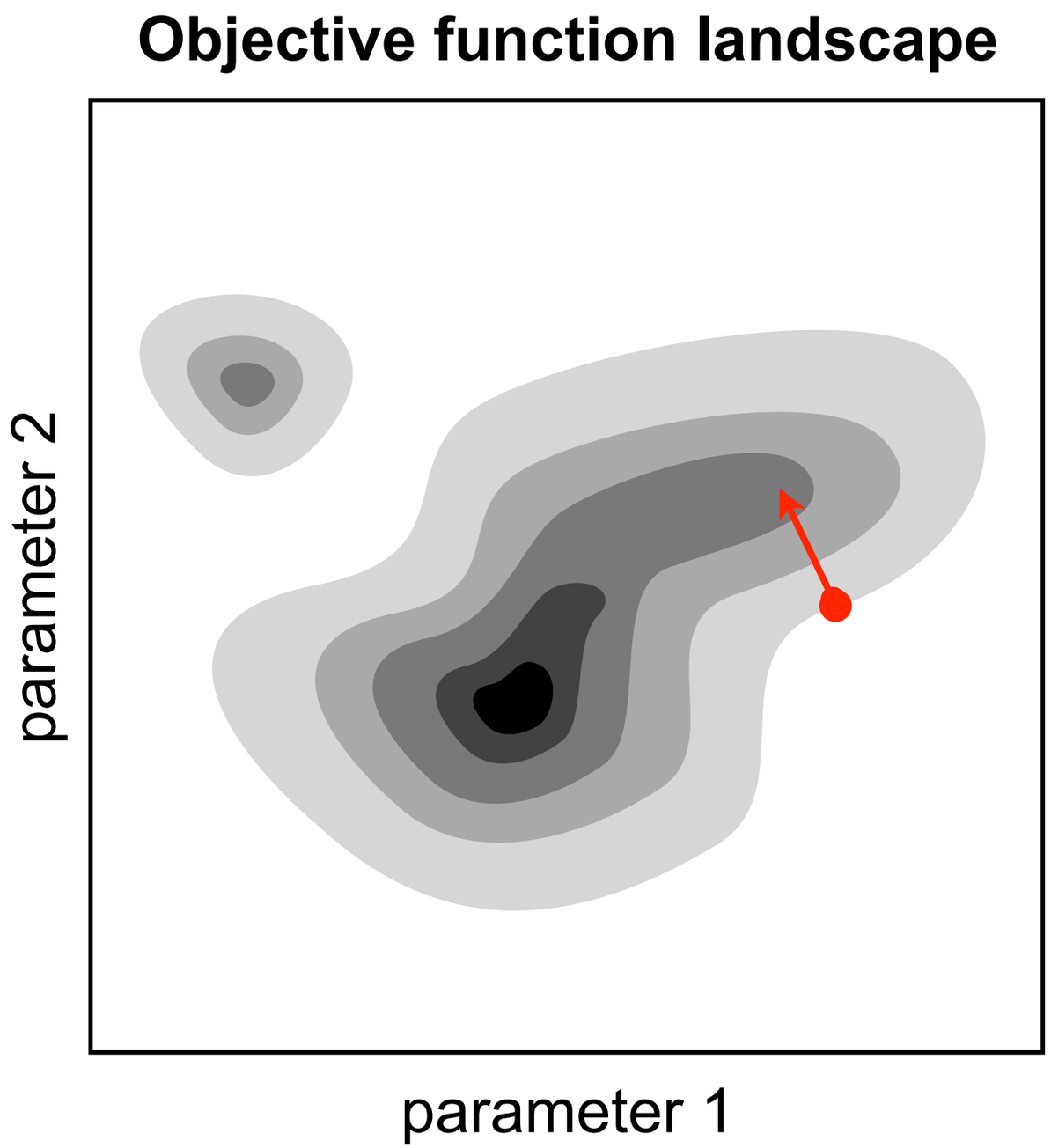
Multi-start local optimisation



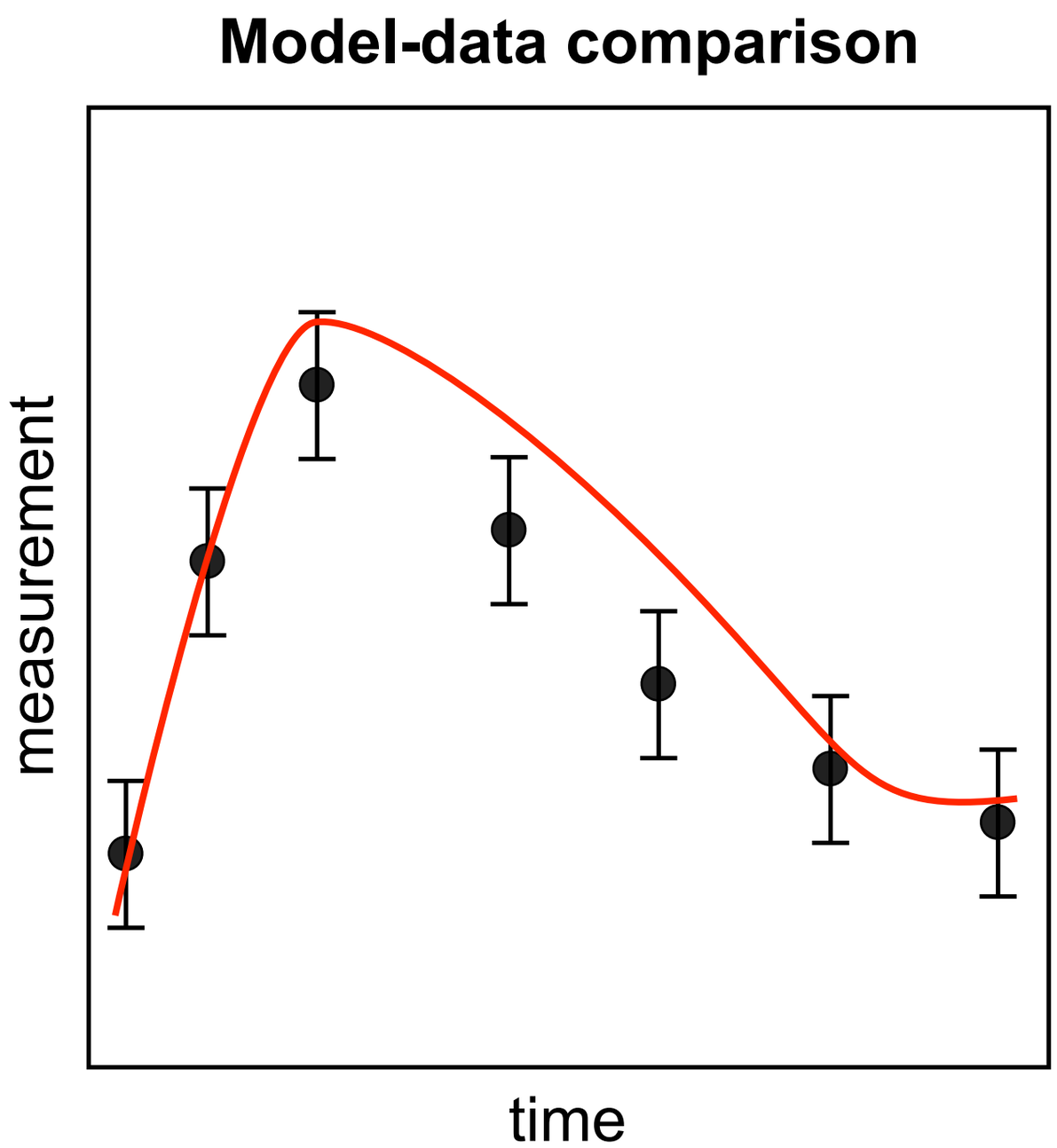


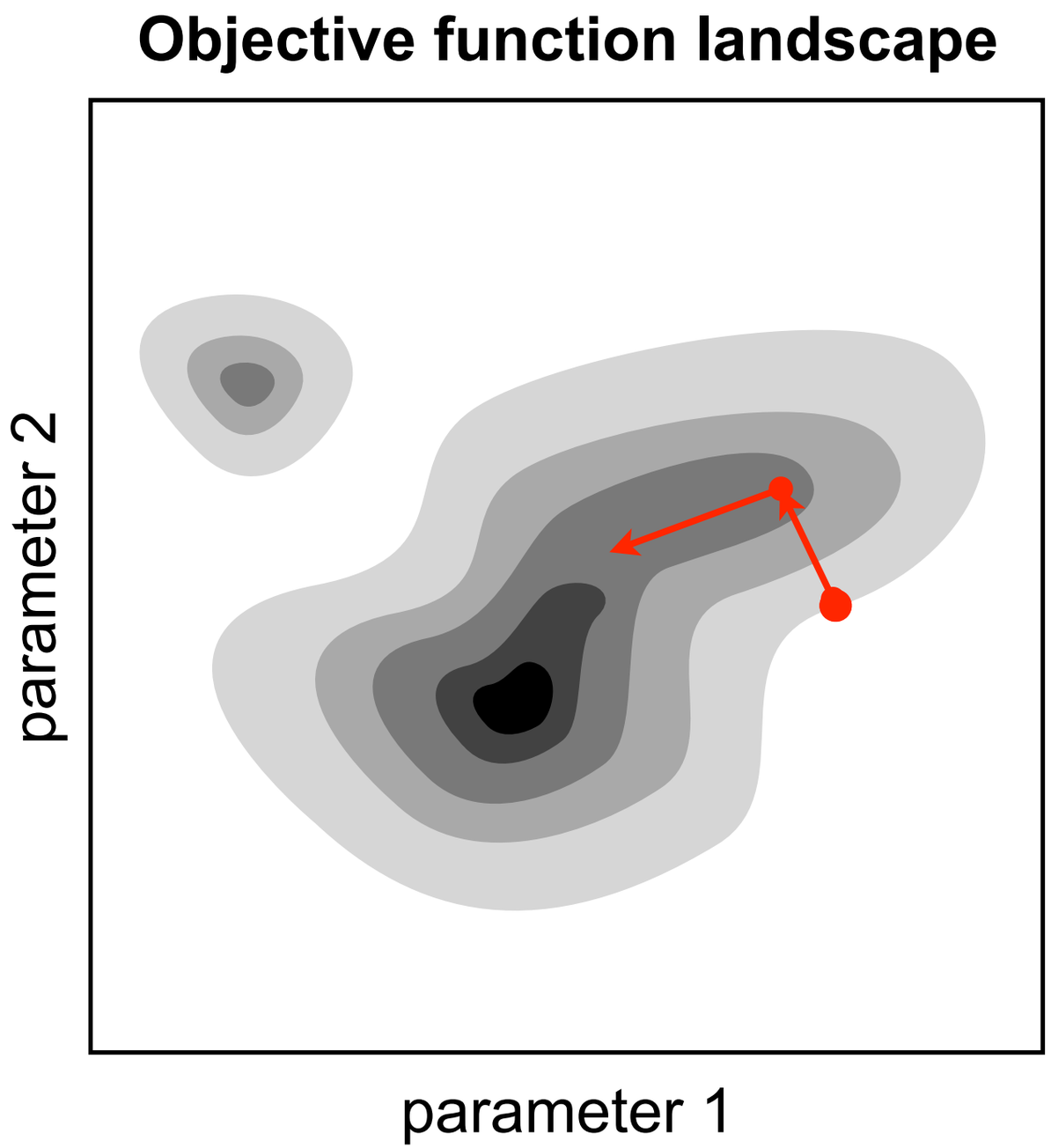
Multi-start local optimisation



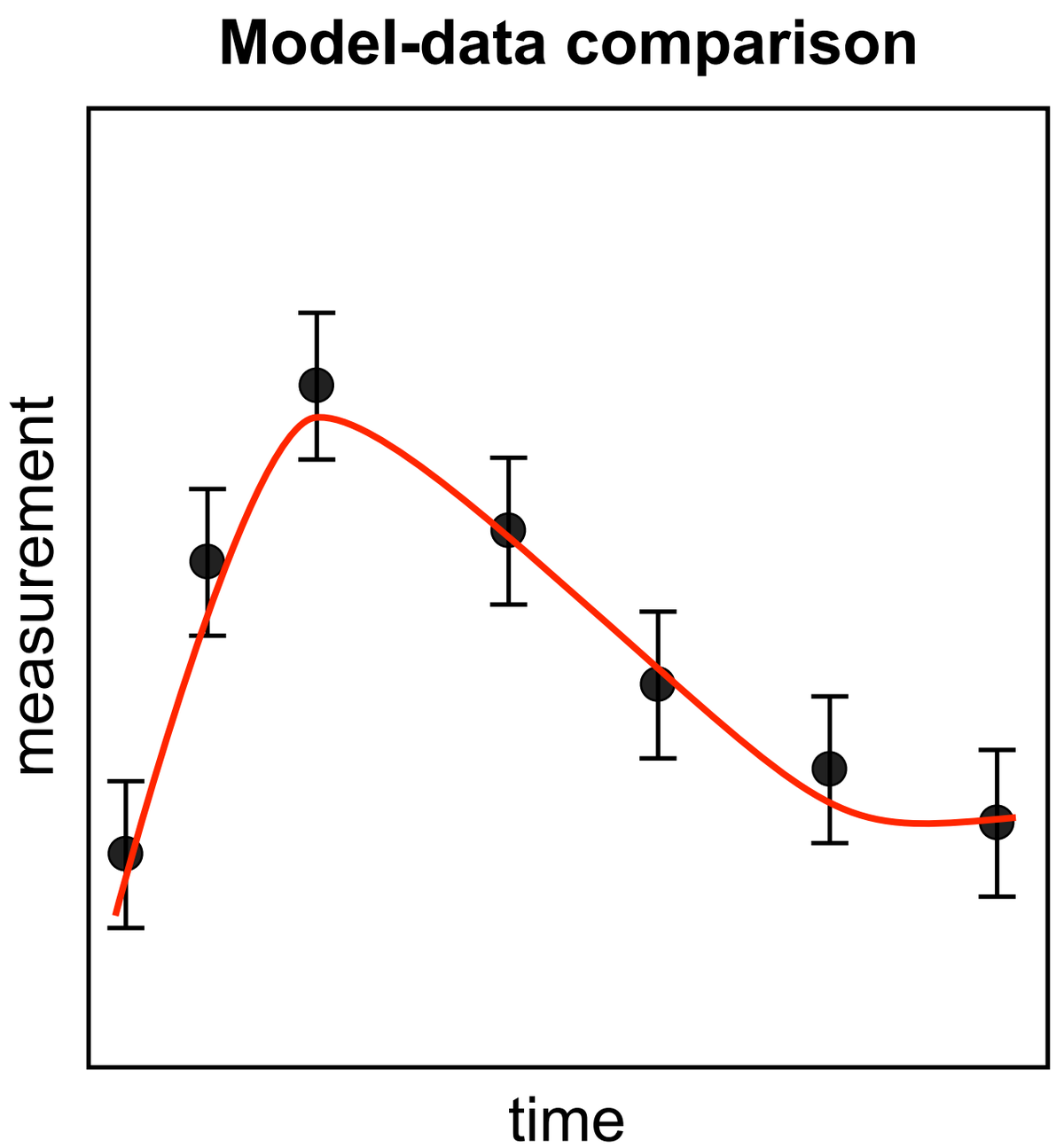


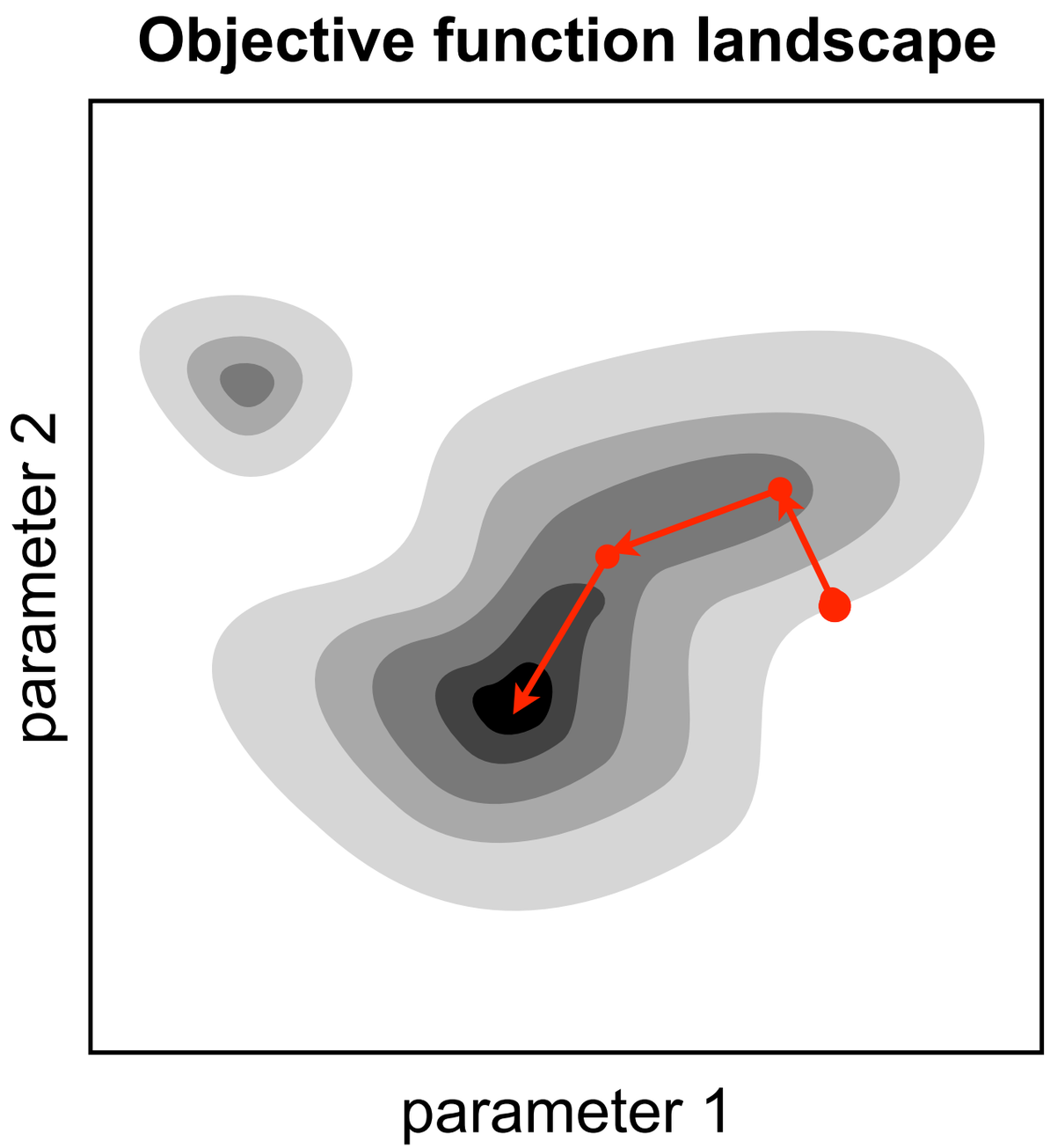
Multi-start local optimisation



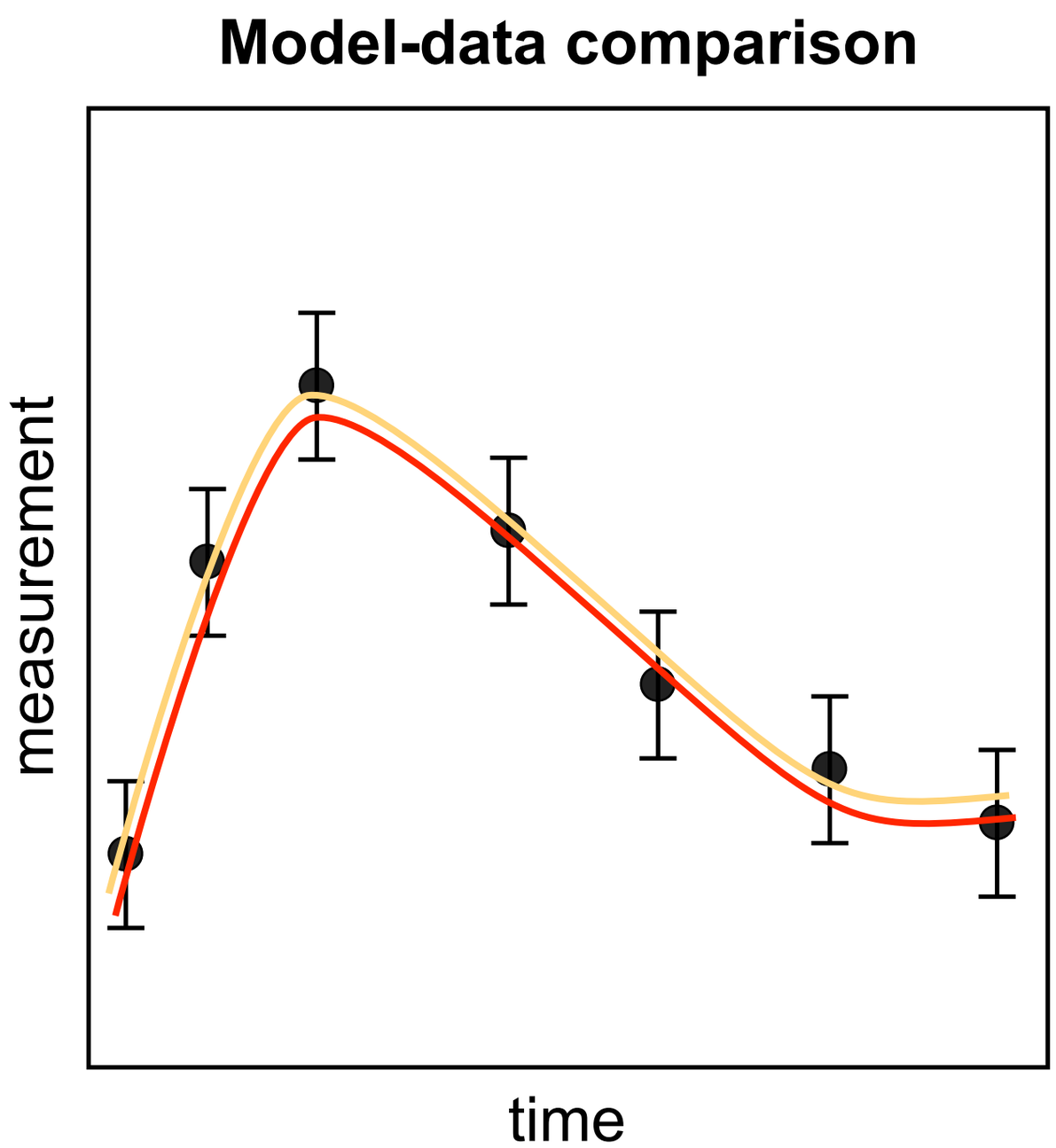


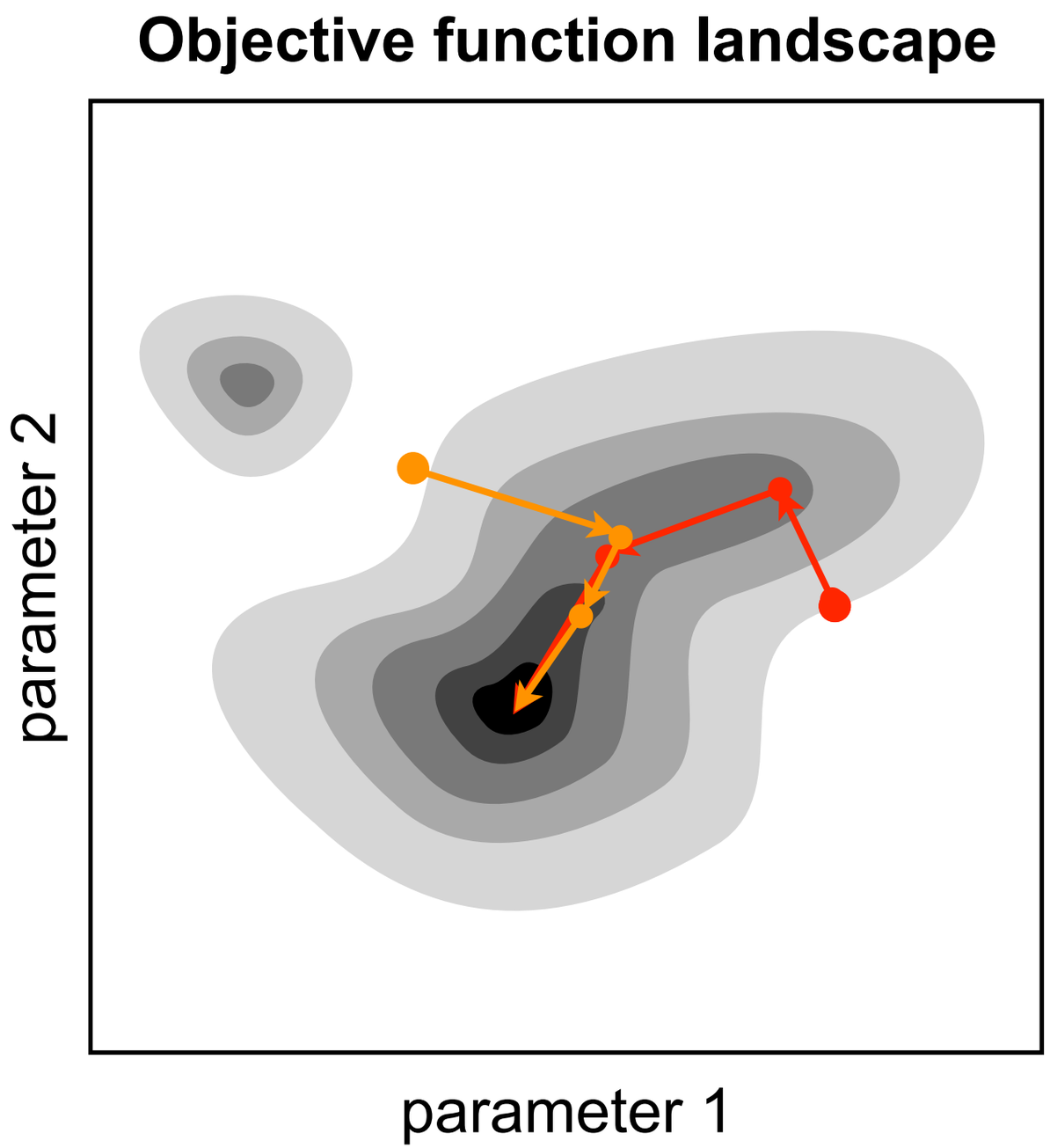
Multi-start local optimisation



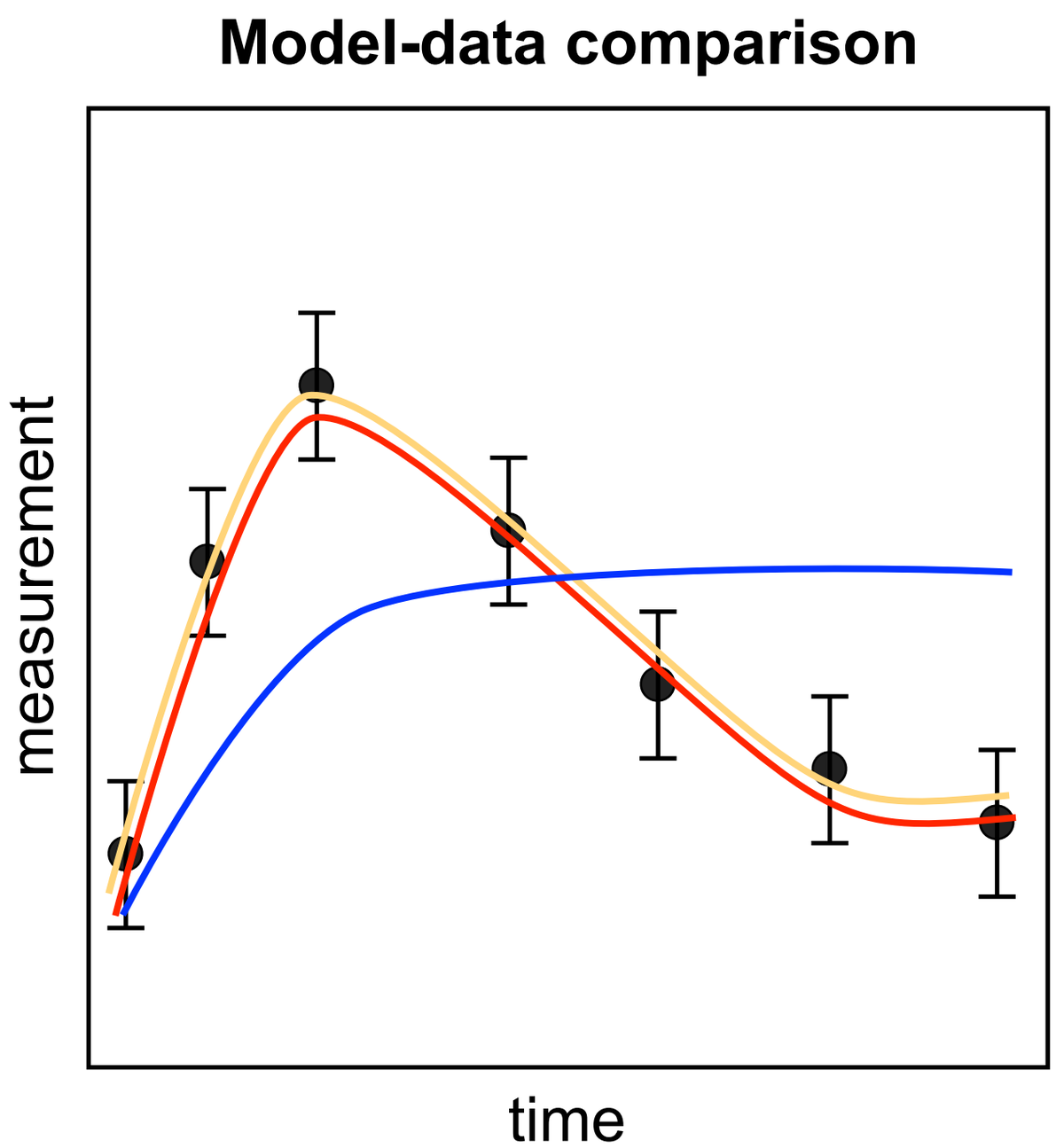


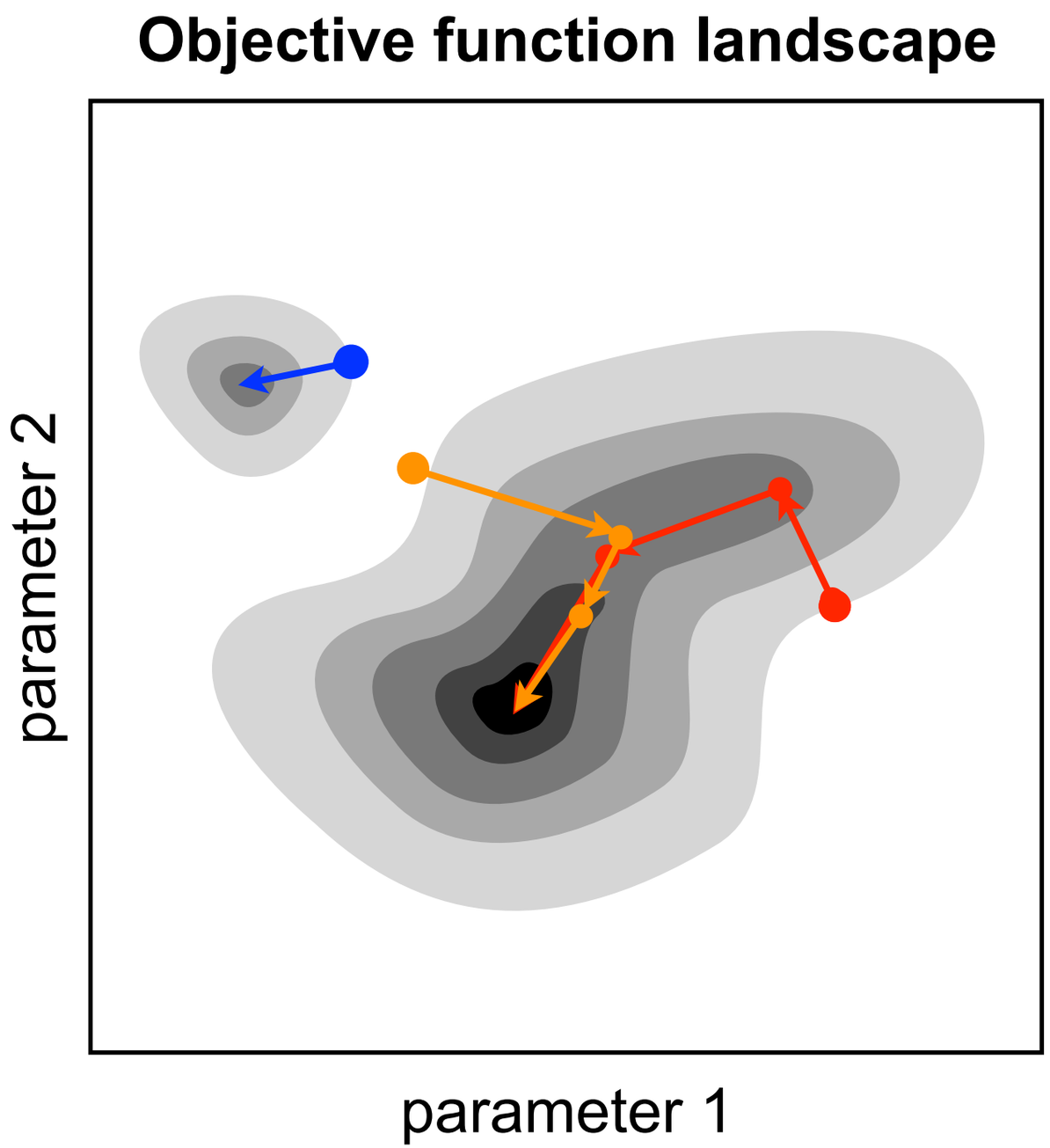
Multi-start local optimisation



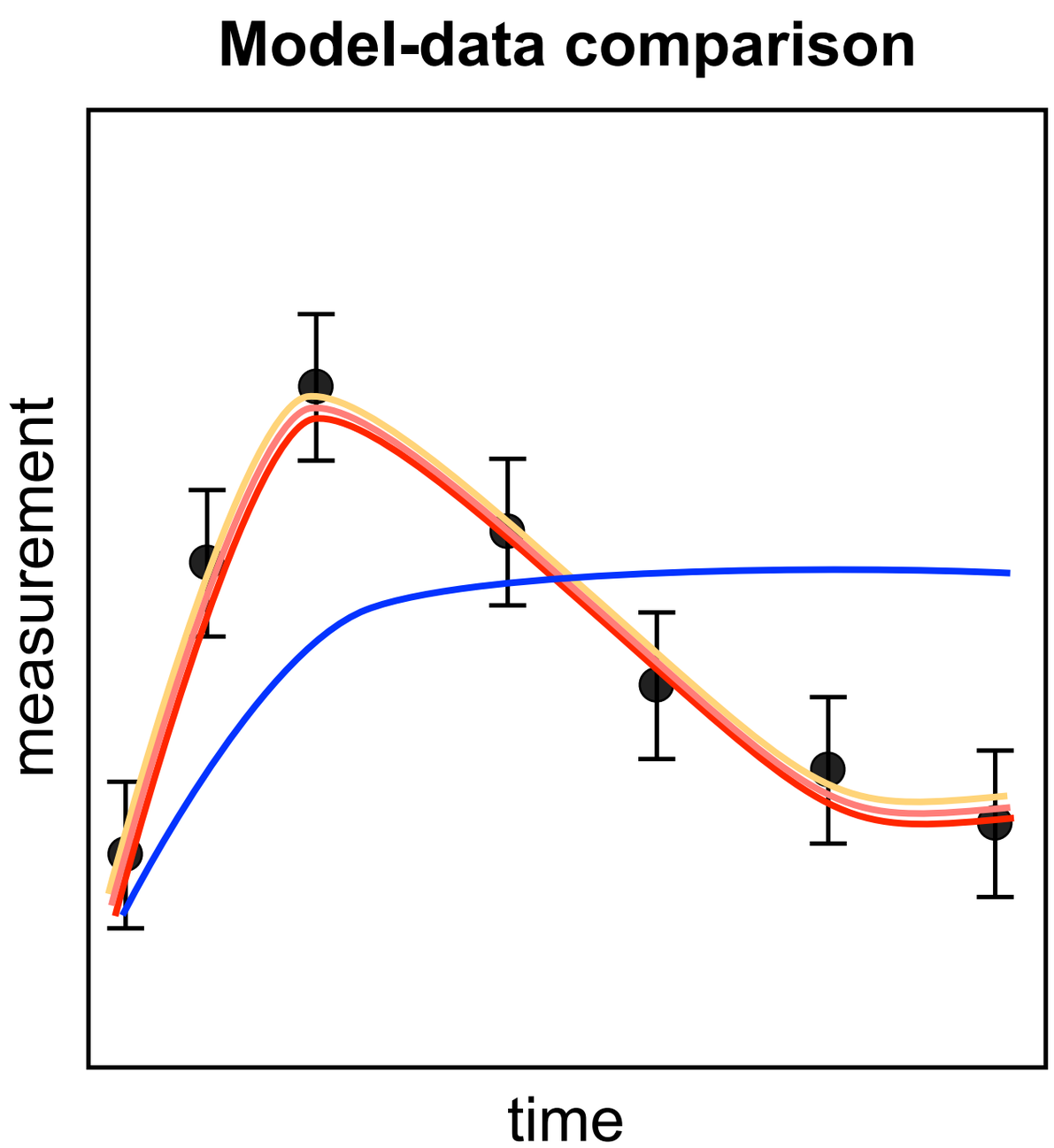


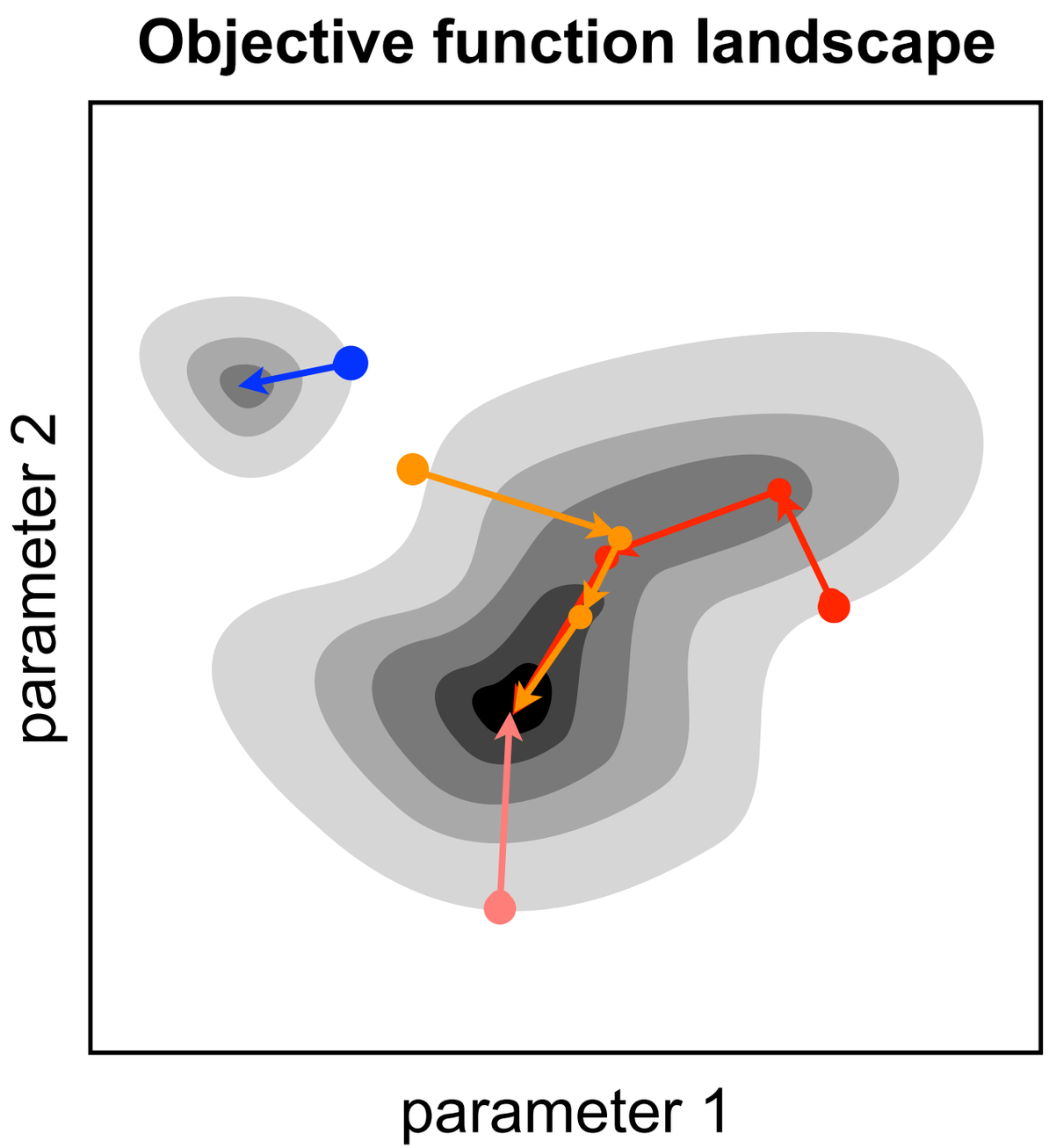
Multi-start local optimisation



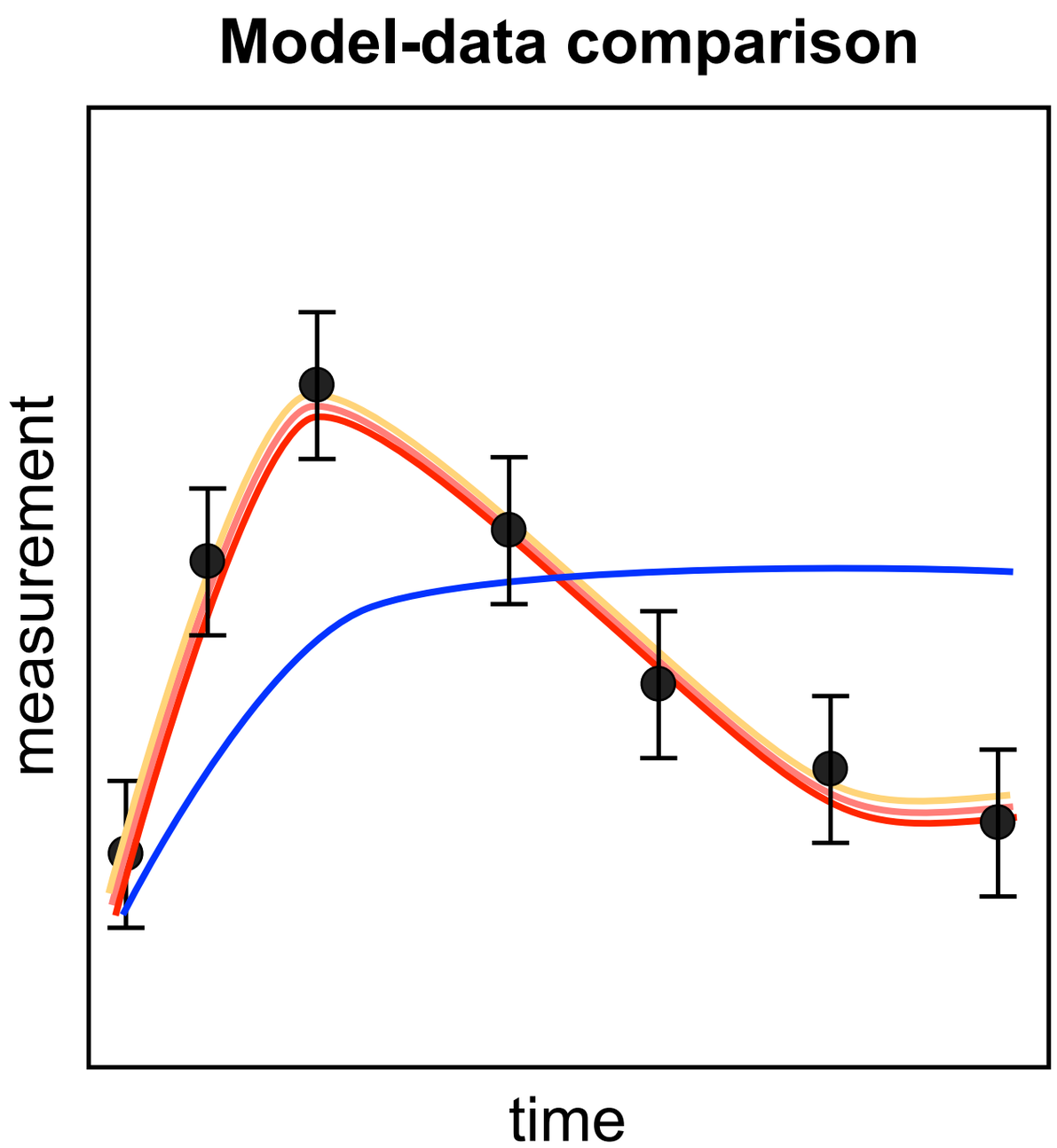


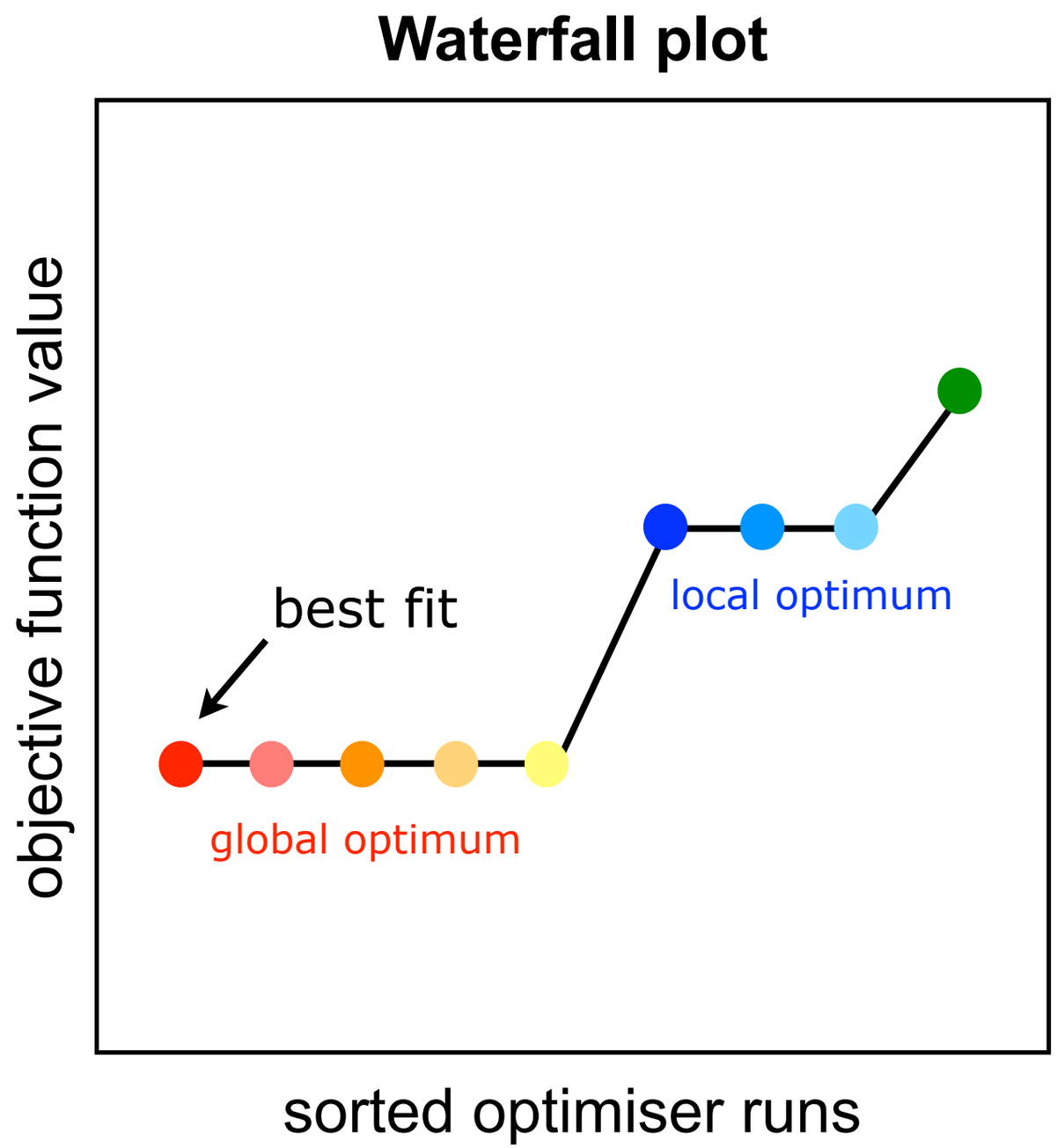
Multi-start local optimisation



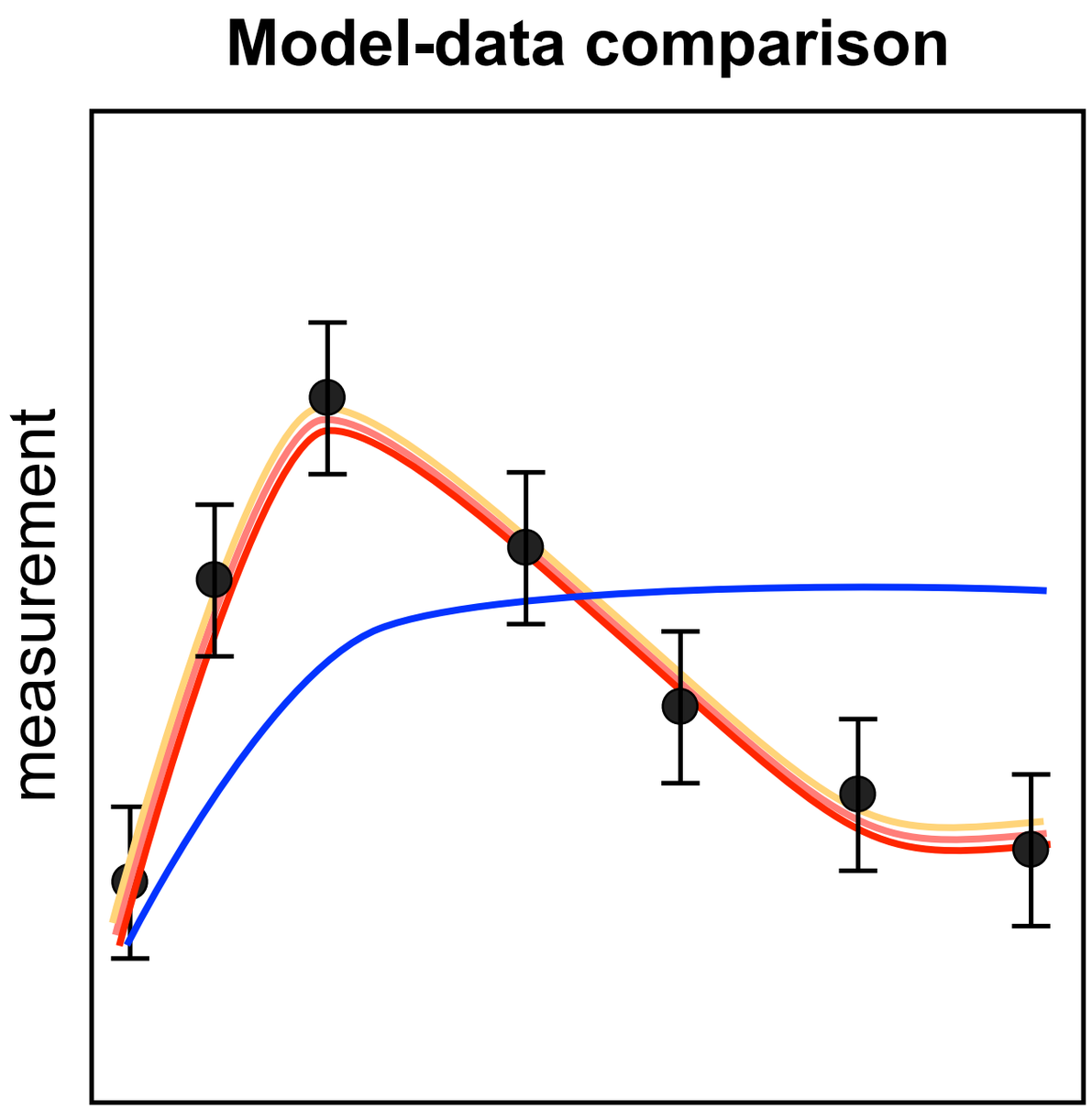


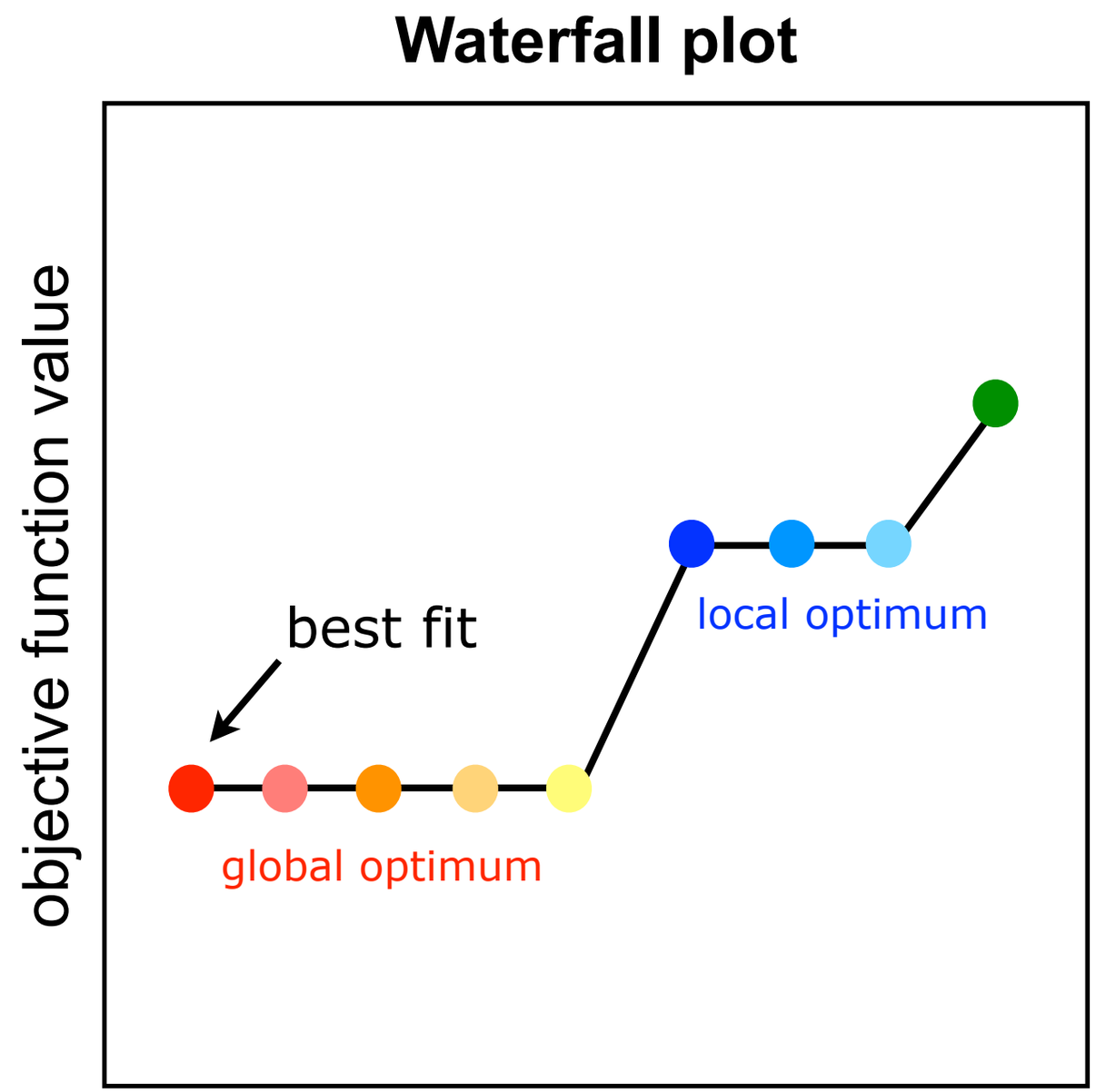
Multi-start local optimisation





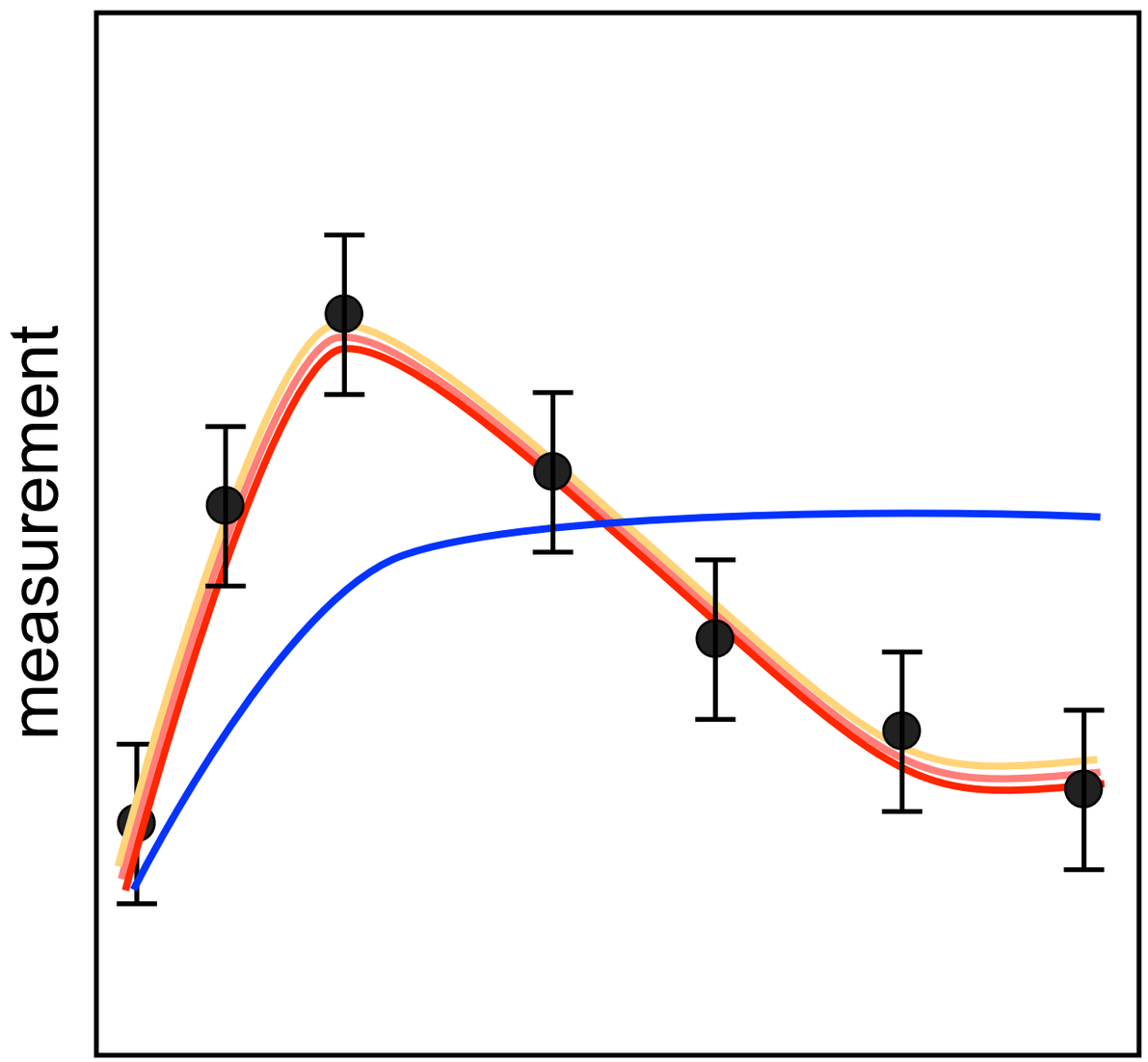
Multi-start local optimisation

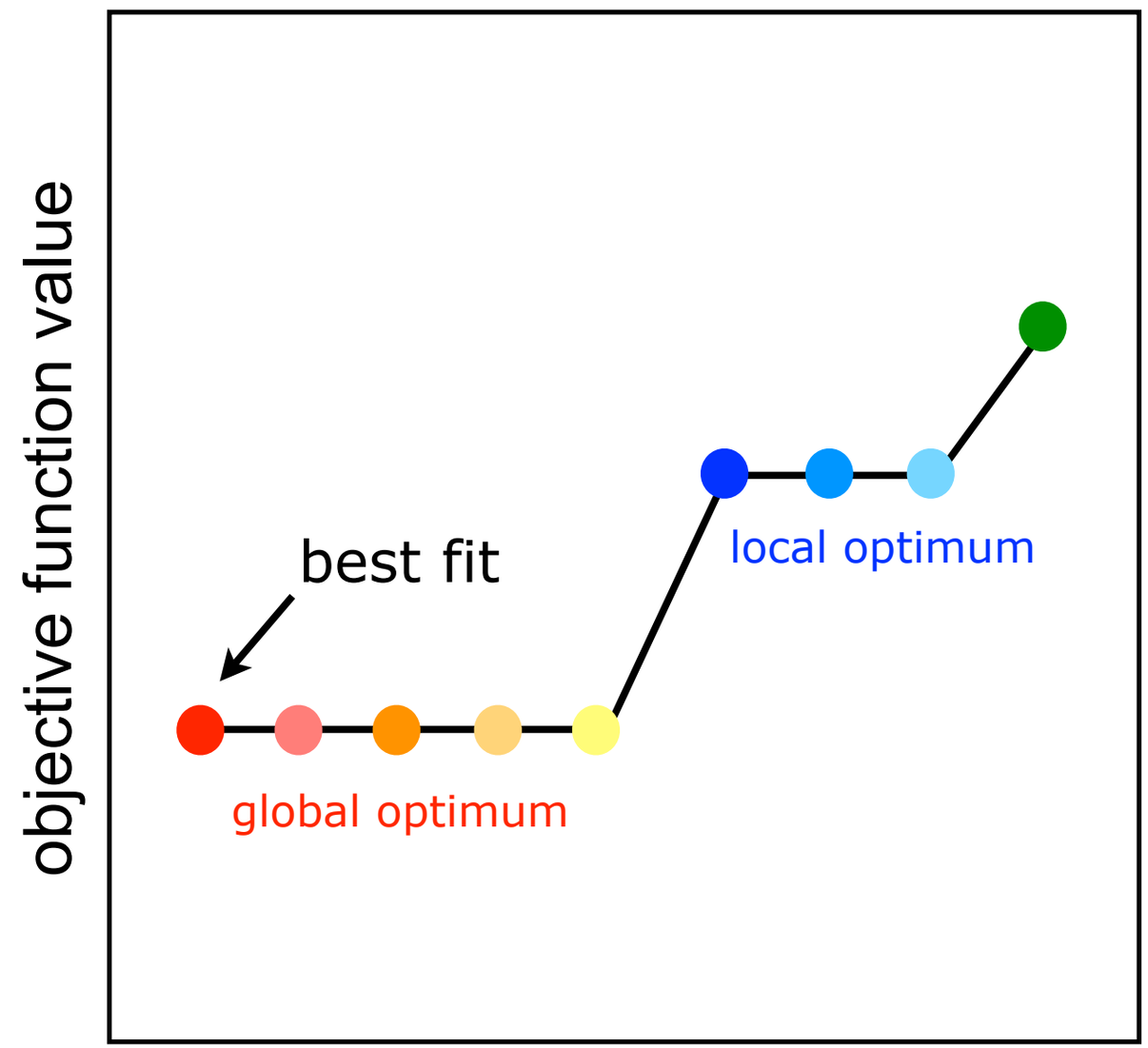




Always check the reproducibility of the fitting results.

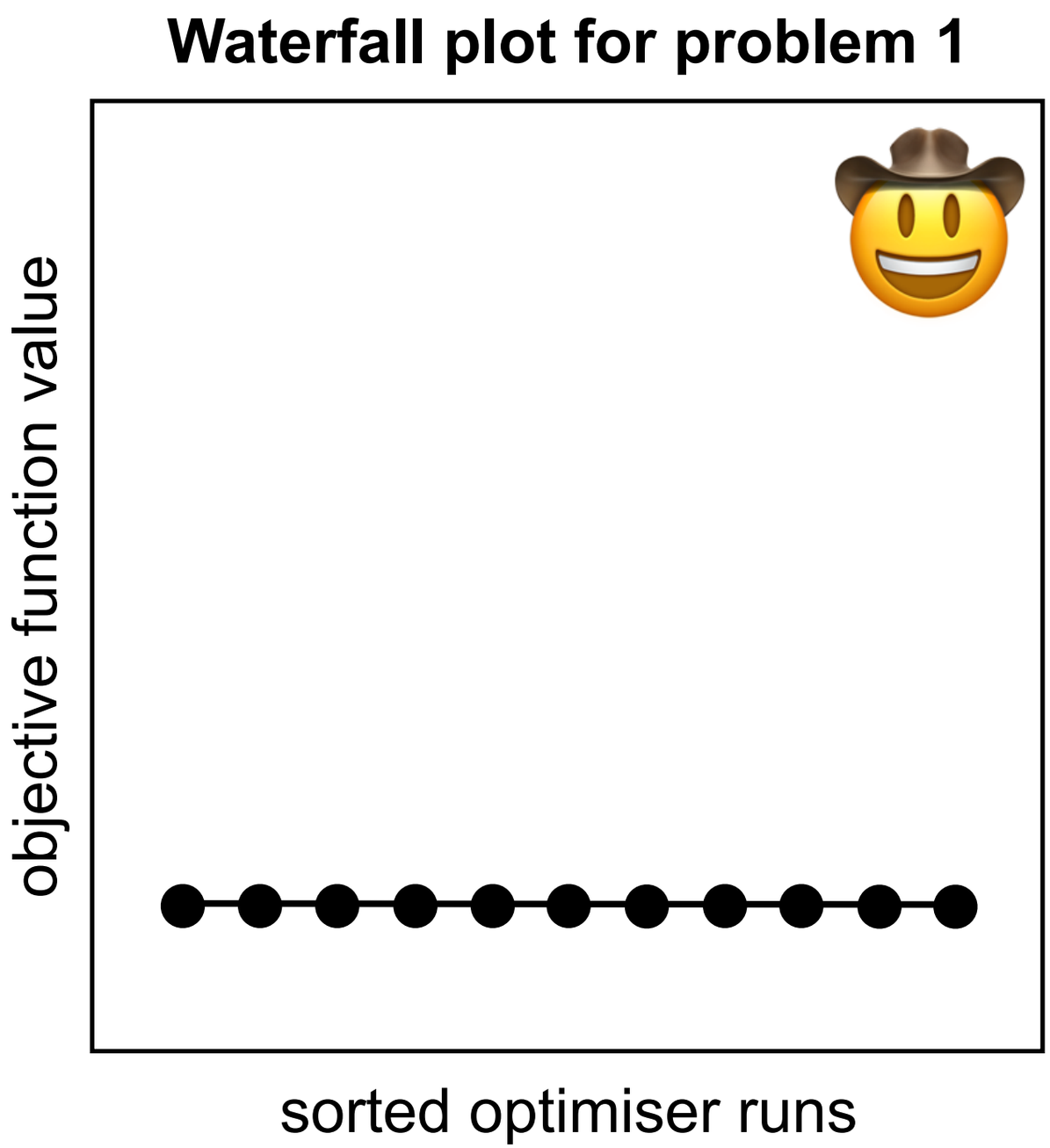
Multi-start local optimisation



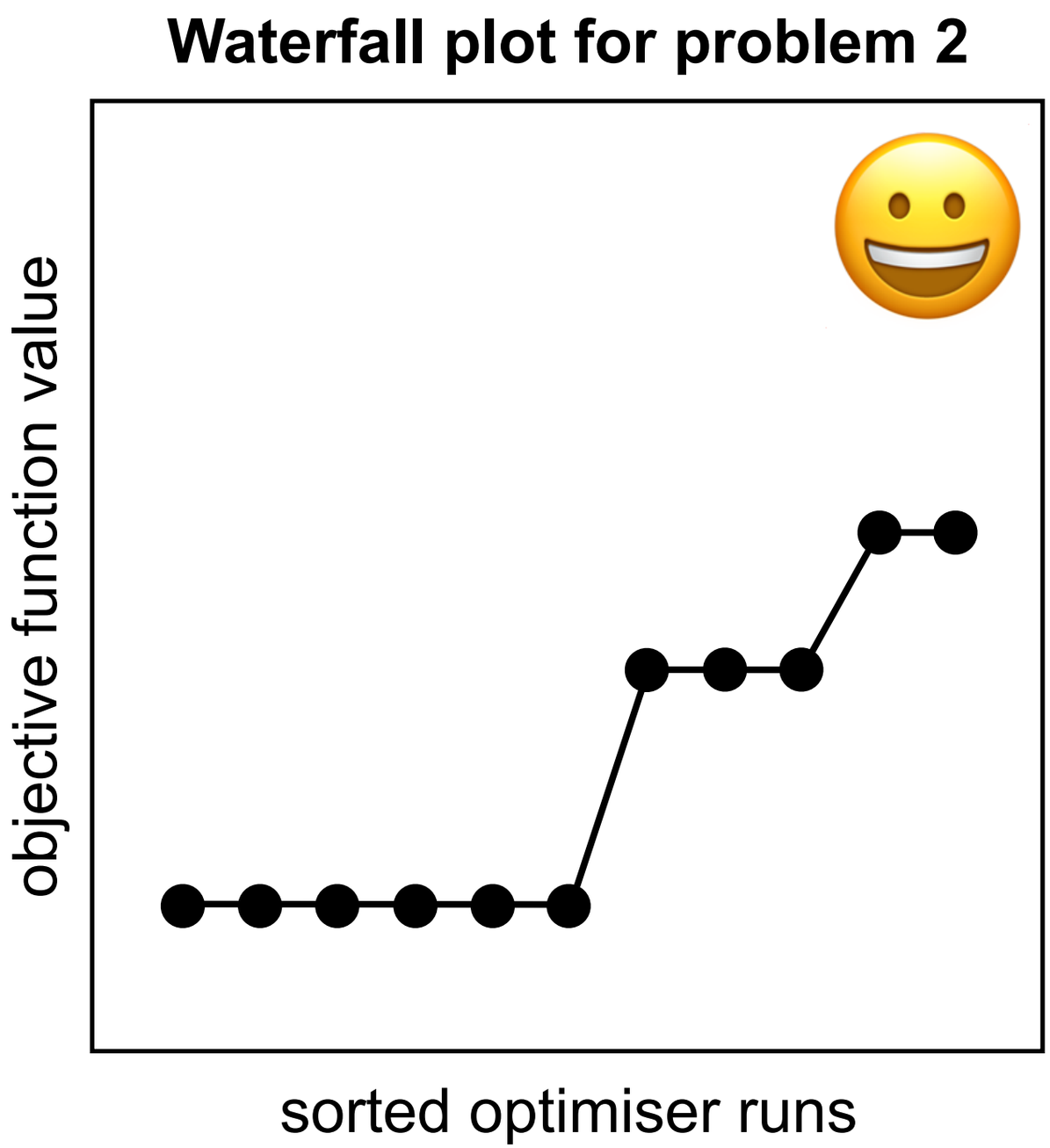


… also if you use global optimisation methods!

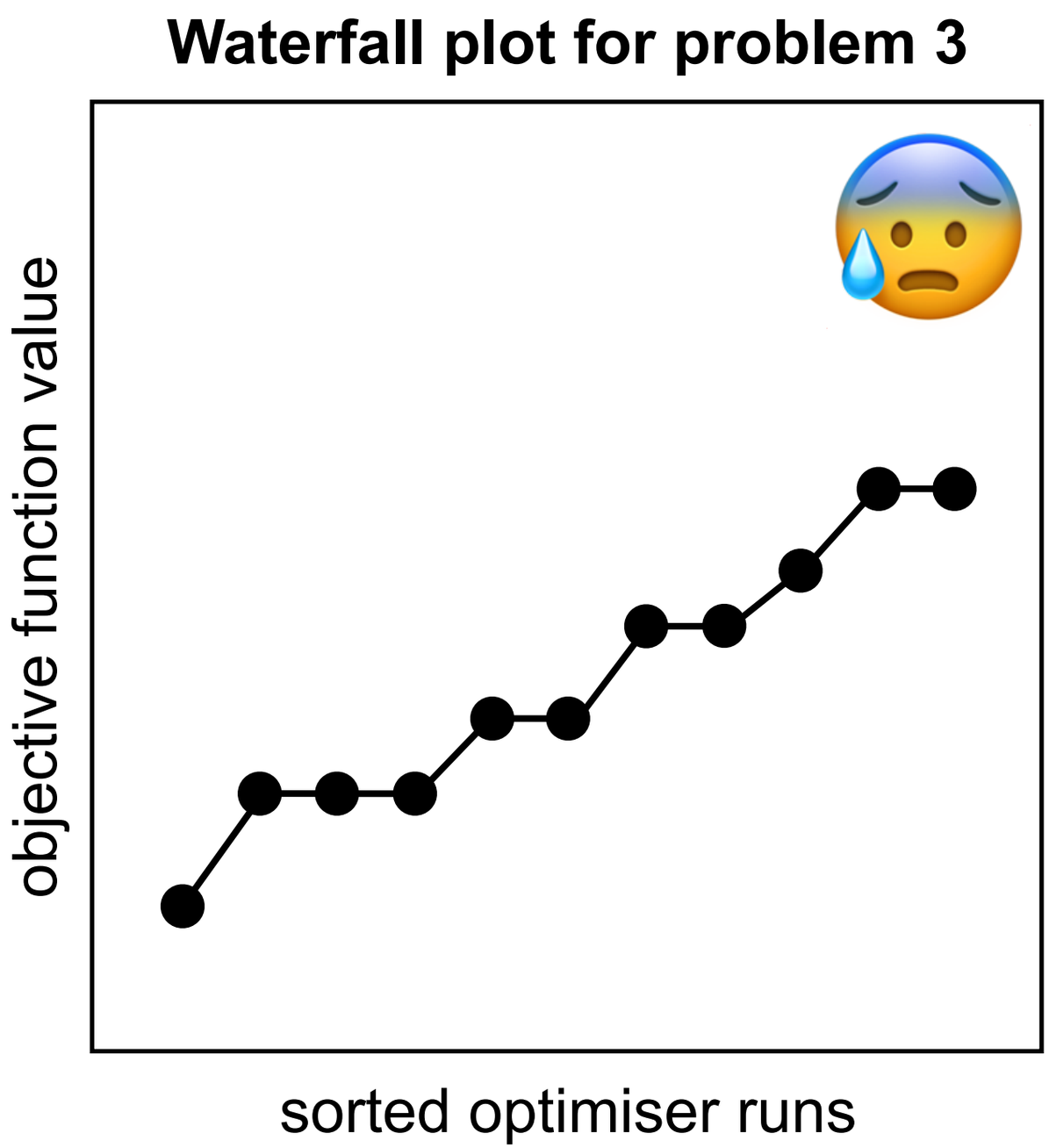
Interpretation of waterfall plots



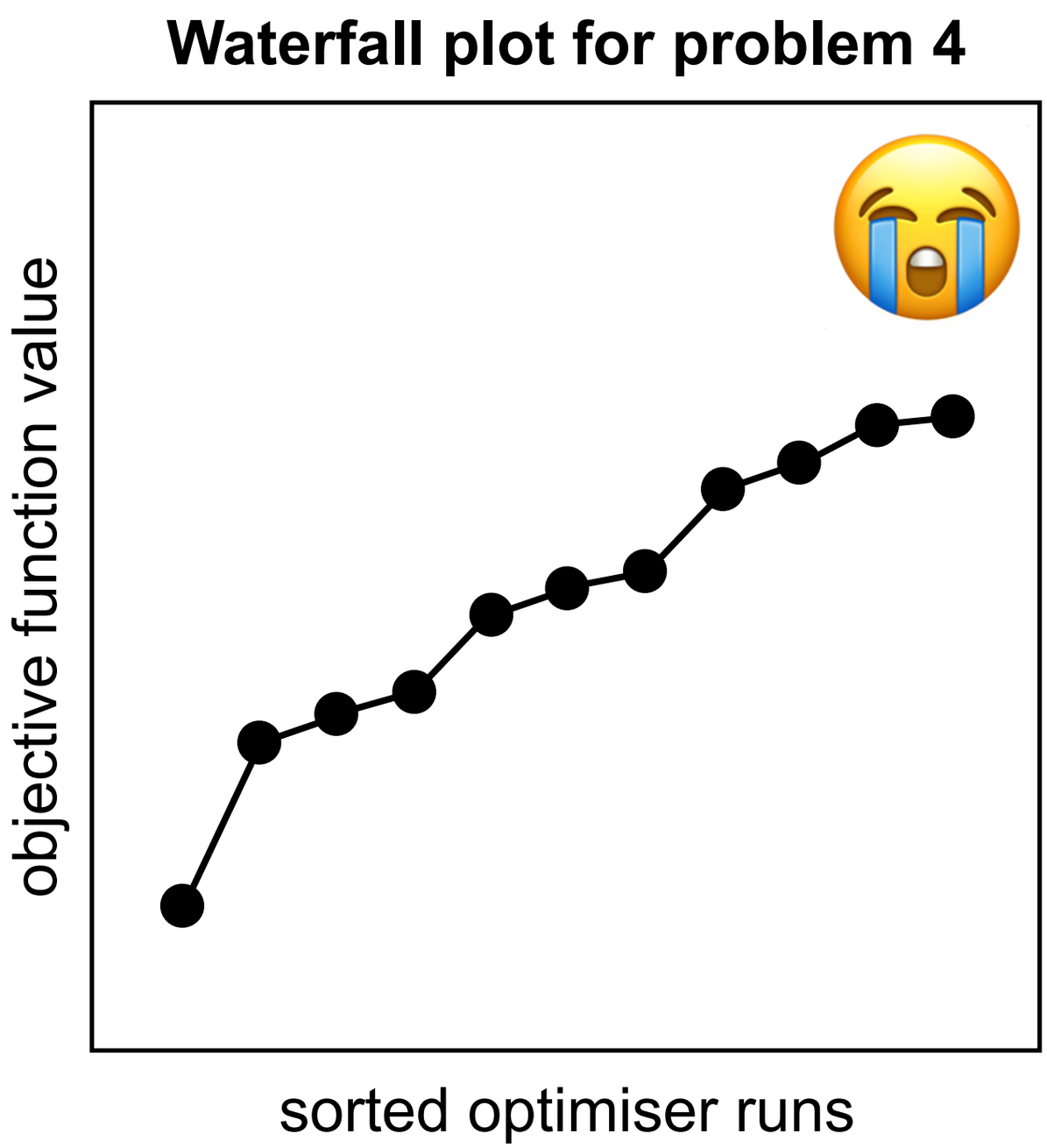
Interpretation of waterfall plots



Interpretation of waterfall plots



Interpretation of waterfall plots



Global optimisation using alternative methods

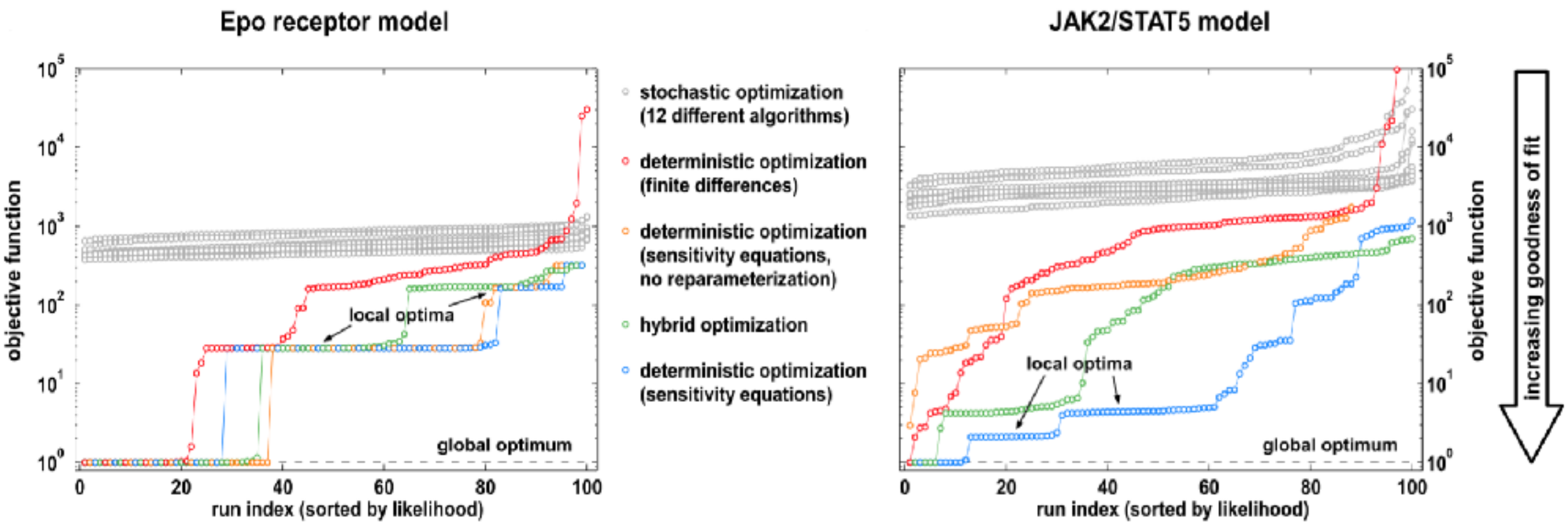
Alternative global optimisation algorithms

- Deterministic methods
  - branch-and-bound
  - interval optimisation
- Stochastic, thermodynamic methods
  - simulated annealing
  - evolutionary algorithms
  - swarm-based optimisation algorithms
- Hybrid stochastic-deterministic methods

Common claim:
Many global optimisation procedure better than multi-start optimisation.

Personal experience: Multi-start local optimisation is for medium- and large-scale optimisation problems in systems / computational biology often better than other methods, given that the local optimiser works well.

Comparison of global optimisation problems

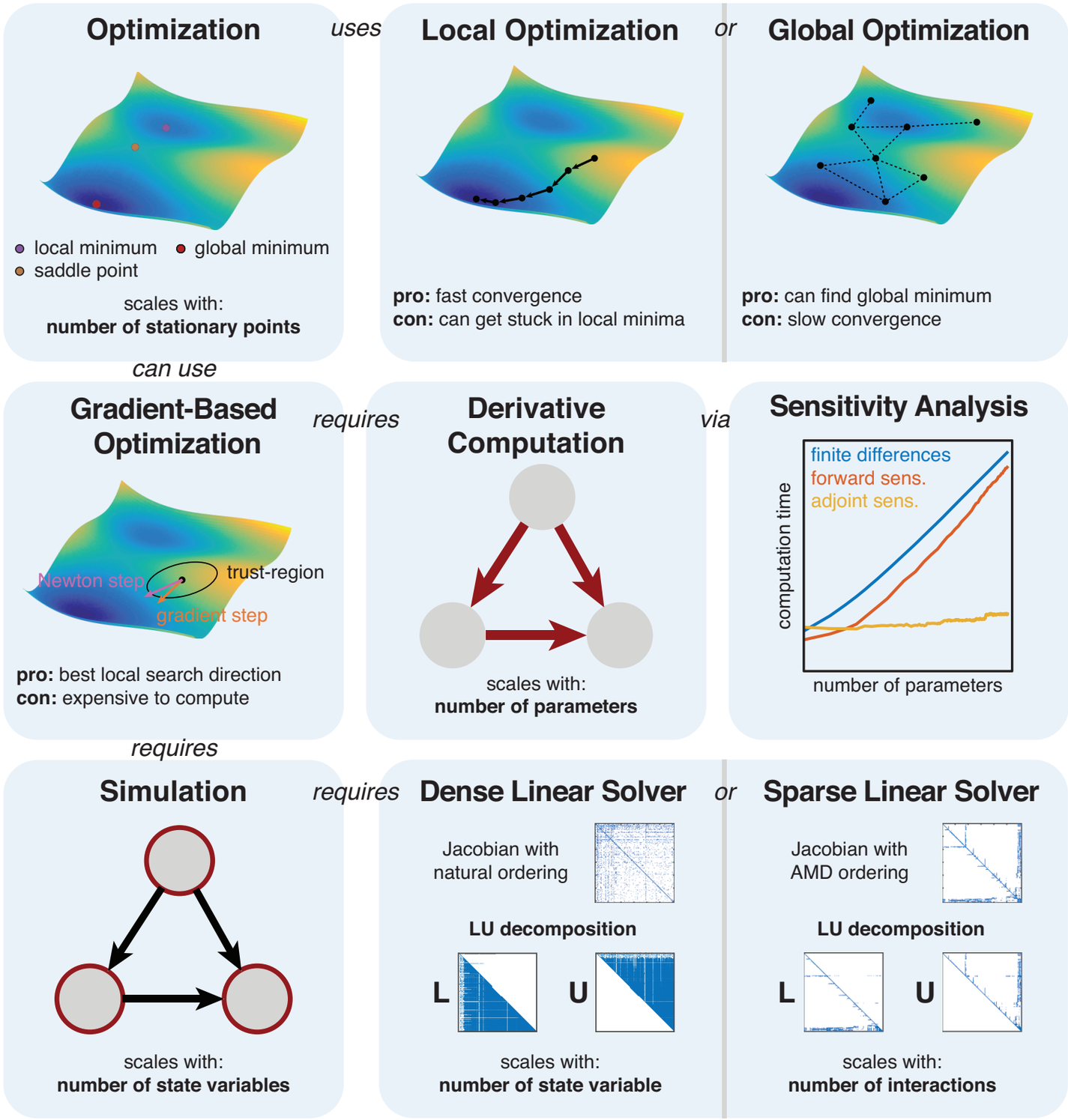


Observation 1: Only multi-start local and hybrid optimisation methods achieves convergence for high-dimensional problems.

Observation 2: Re-parameterisation (i.e., log-transformation) is not only important for the efficiency but also the convergence of optimisers.

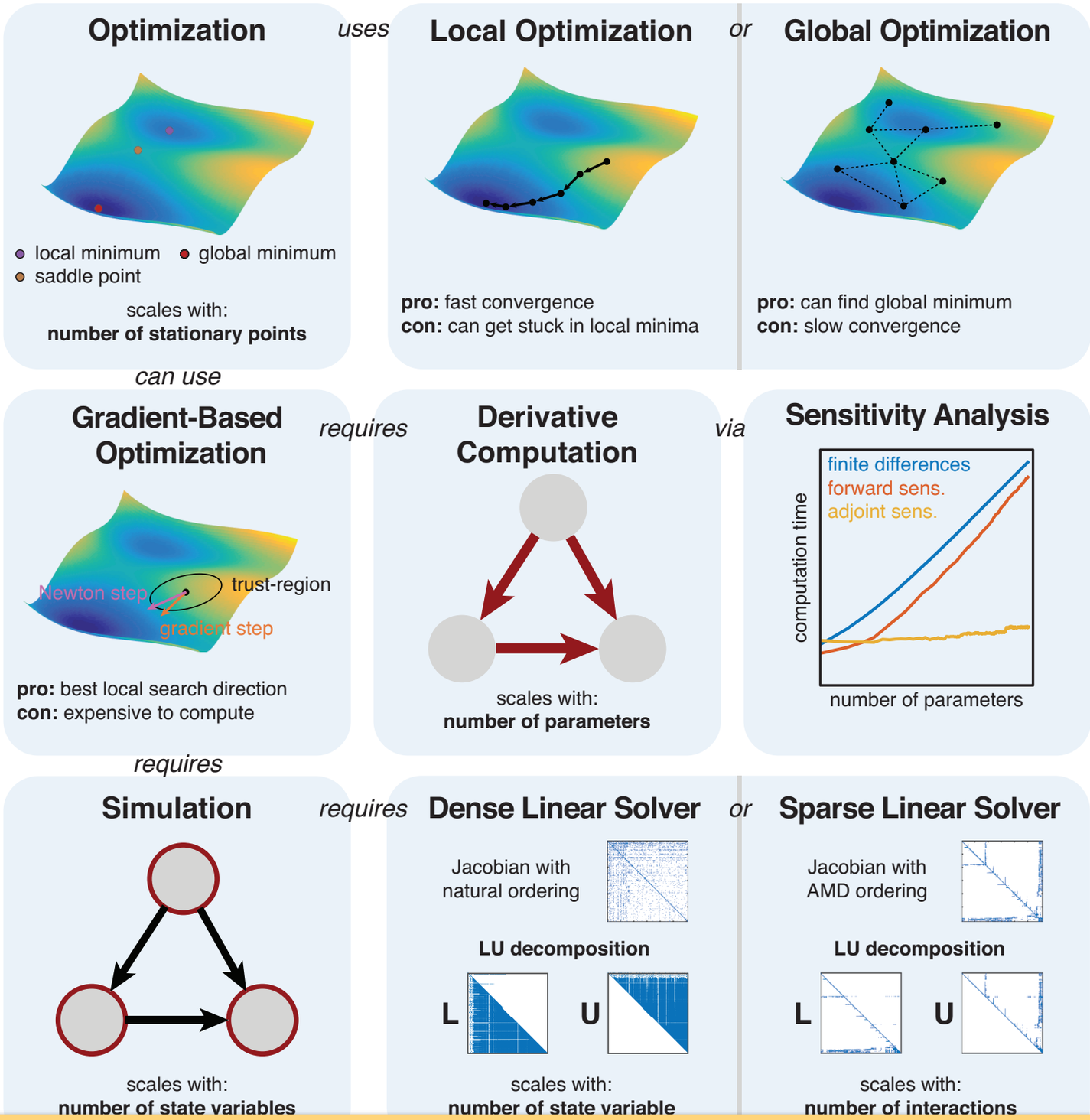
Raue et al., Lessons learned from quantitative dynamical modeling in systems biology, PLoS ONE, 8(9):e74335, 2013.

Ingredients of a good optimisation method



F. Fröhlich, C. Loos, and J. Hasenauer. Scalable inference of ordinary differential equation models of biochemical processes. In: Gene Regulatory Networks: Methods and Protocols, Humana Press, p. 385-422, 2019.

Ingredients of a good optimisation method



All the steps matter!

Literature

Methods and evaluations

- A. Raue et al. Lessons learned from quantitative dynamical modeling in systems biology. PLoS ONE, 8(9):e74335, 2013.
- J. A. Egea et al. MEIGO: An open-source software suite based on metaheuristics for global optimization in systems biology and bioinformatics. BMC Bioinf., 15(136), 2014.
- A. F. Villaverde, F. Fröhlich, D. Weindl, J. Hasenauer, and J. R. Banga, Benchmarking optimization methods for parameter estimation in large kinetic models. Bioinformatics, 2019.
- H. Hass, C. Loos, E. Raimundez Alvarez, J. Timmer, J. Hasenauer, and C. Kreutz, Benchmark Problems for Dynamic Modeling of Intracellular Processes. Bioinformatics, 2019.

Comprehensive review

- F. Fröhlich, C. Loos, and J. Hasenauer. Scalable inference of ordinary differential equation models of biochemical processes. In: Gene Regulatory Networks: Methods and Protocols, Humana Press, p. 385-422, 2019.

No free lunch theorem

- D. H. Wolpert and W. G. Macready. No free lunch theorems for optimization. IEEE Trans. Evol. Comput., 1(1):67-82,1997.
- A. Auger and O. Teytaud. Continuous lunches are free plus the design of optimal optimization algorithms. Algorithmica, 57(1):121-146, 2010.

How can we deal with relative data?

Parameter estimation for relative data

ODE model:

$$\frac{dx}{dt} = f(\theta, x(t,\theta)), \quad x(0,\theta) = x_0(\theta) \quad \textit{dynamics}$$

$$y(t) = c \cdot h(\theta, x(t,\theta)) \quad \textit{observables}$$

Measurements that provide relative data:

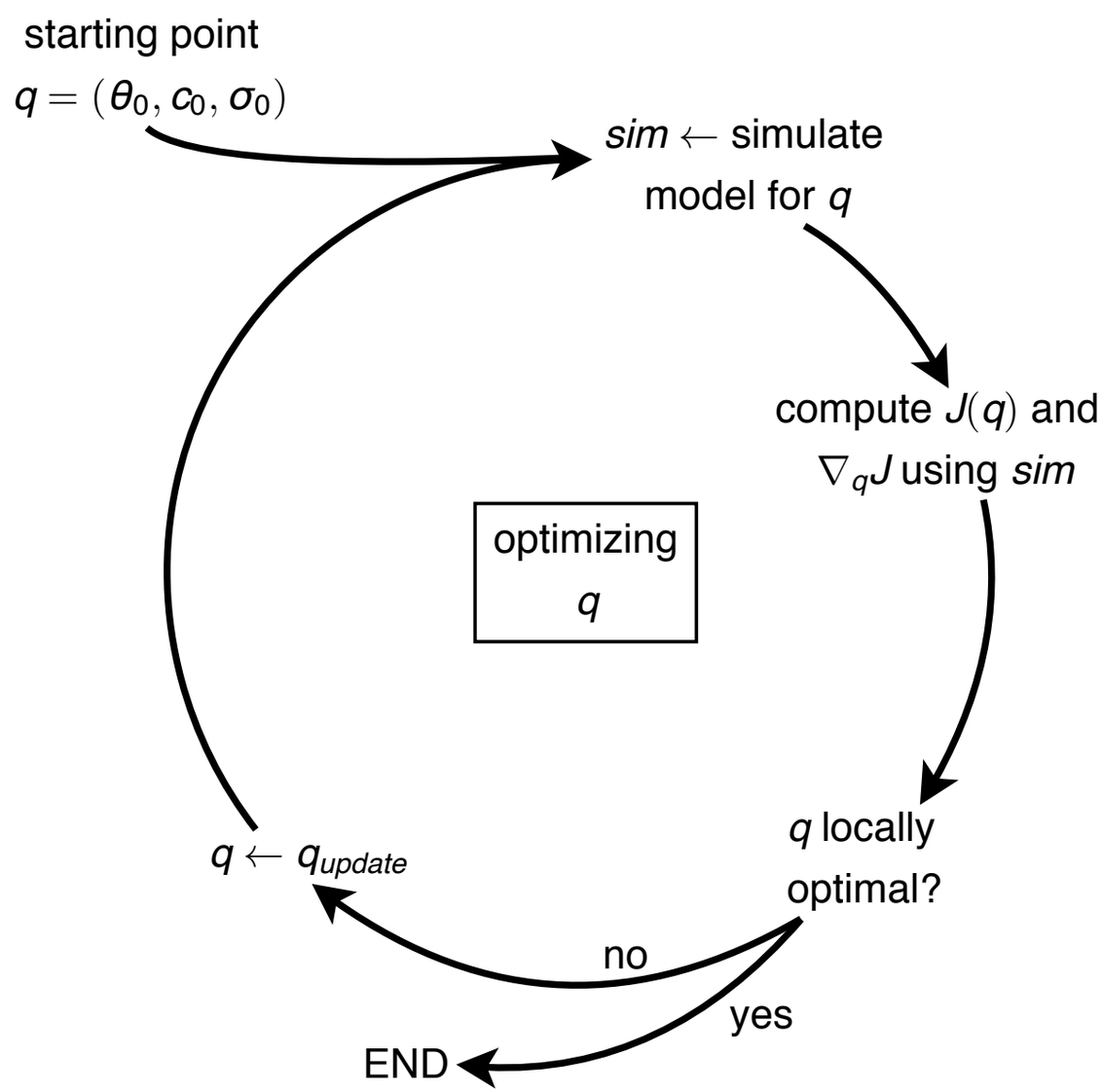
$$\bar{y}_k = c \cdot h(\theta, x(t_k,\theta)) + \varepsilon_k, \quad \varepsilon_k \sim \mathcal{N}(0,\sigma^2), \quad k = 1,\ldots,n_t$$

with unknown variance σ^2 of the measurement noise
and unknown proportionality factor c

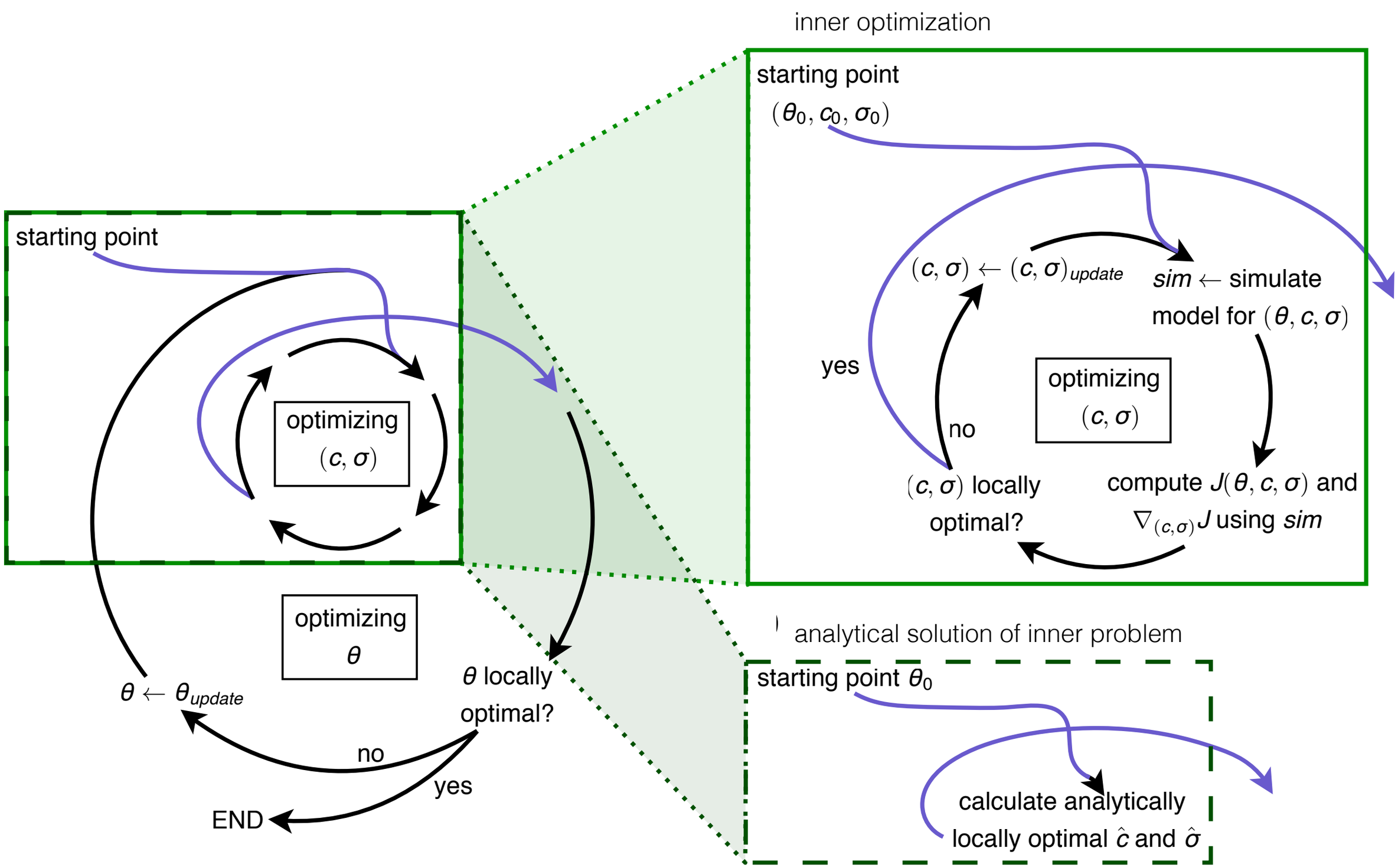
Minimise the negative log likelihood function:

$$\min_{(\theta,c,\sigma)} \left\{ J(\theta,c,\sigma^2) = \frac{1}{2}\sum_k \log(2\pi\sigma^2) + \left(\frac{\bar{y}_k - c \cdot h(\theta, x(t_k,\theta))}{\sigma} \right)^2 \right\}$$

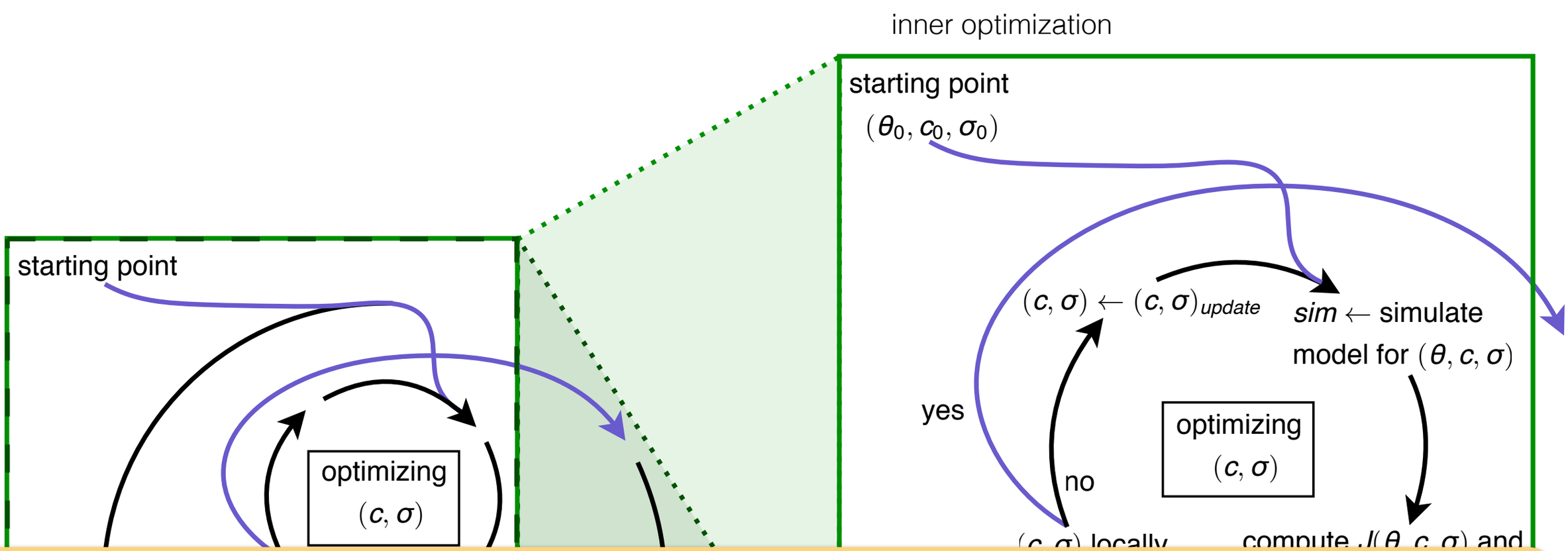
Standard optimisation approach



Hierarchical optimisation approach



Hierarchical optimisation approach

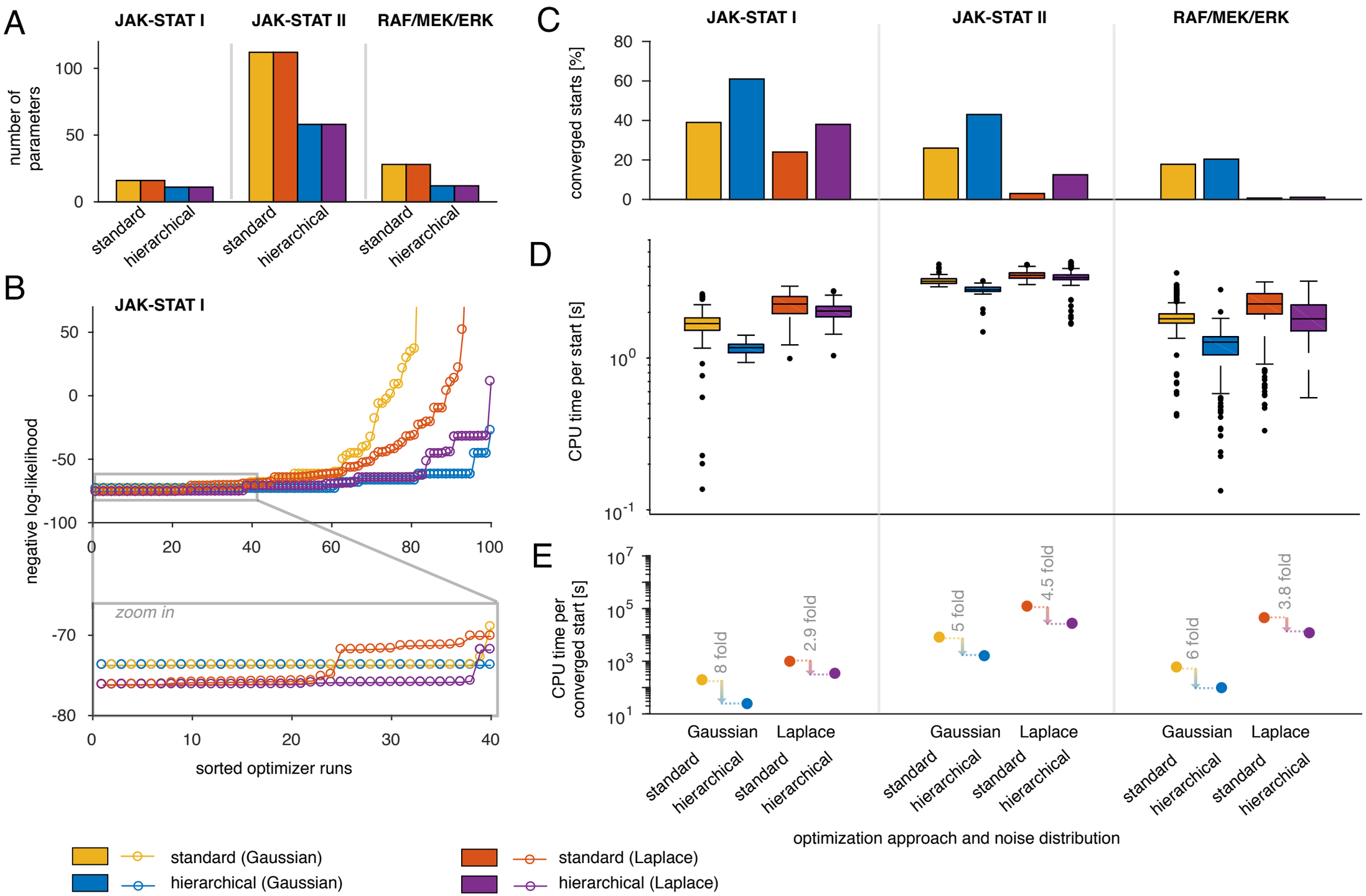


Hierarchical optimisation problem:

$$\min_{(\theta} \left\{ \min_{(c,\sigma)} \left\{ J(\theta, c, \sigma^2) = \frac{1}{2} \sum_k \log(2\pi\sigma^2) + \left(\frac{\bar{y}_k - c \cdot h(\theta, x(t_k, \theta))}{\sigma} \right)^2 \right\} \right\}$$

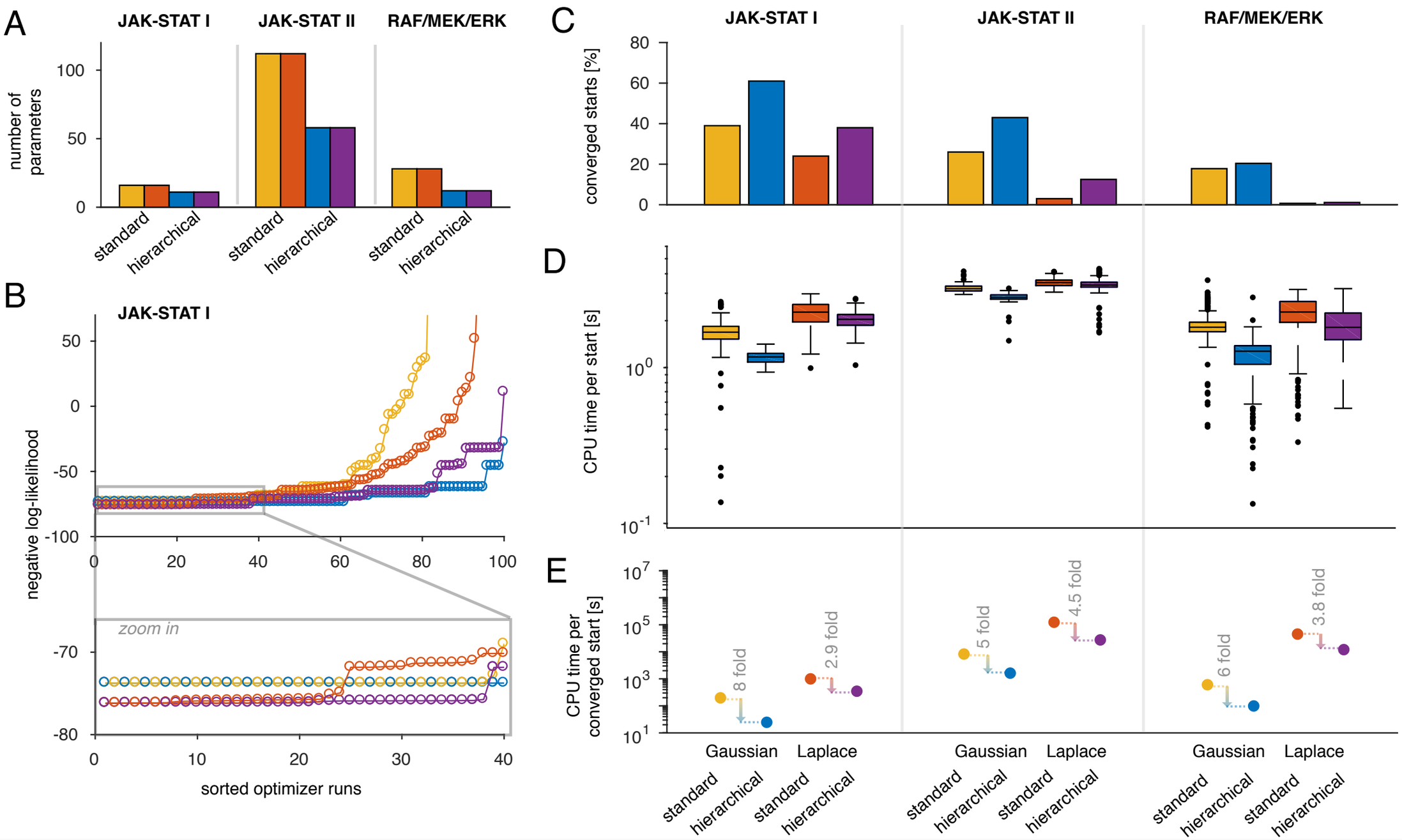
with $\hat{c}(\theta) = \dfrac{\sum_k \bar{y}_k h(\theta, x(t_k, \theta))}{\sum_k h(\theta, x(t_k, \theta))^2}$ and $\hat{\sigma}^2(\theta) = \dfrac{1}{n_t} \sum_k \left(\bar{y}_k - \hat{c} \cdot h(\theta, x(t_k, \theta)) \right)^2$

Evaluation (1)



Loos et al. Hierarchical optimization for the efficient parametrization of ODE models. Bioinformatics, 34(24):4266-4273, 2018.

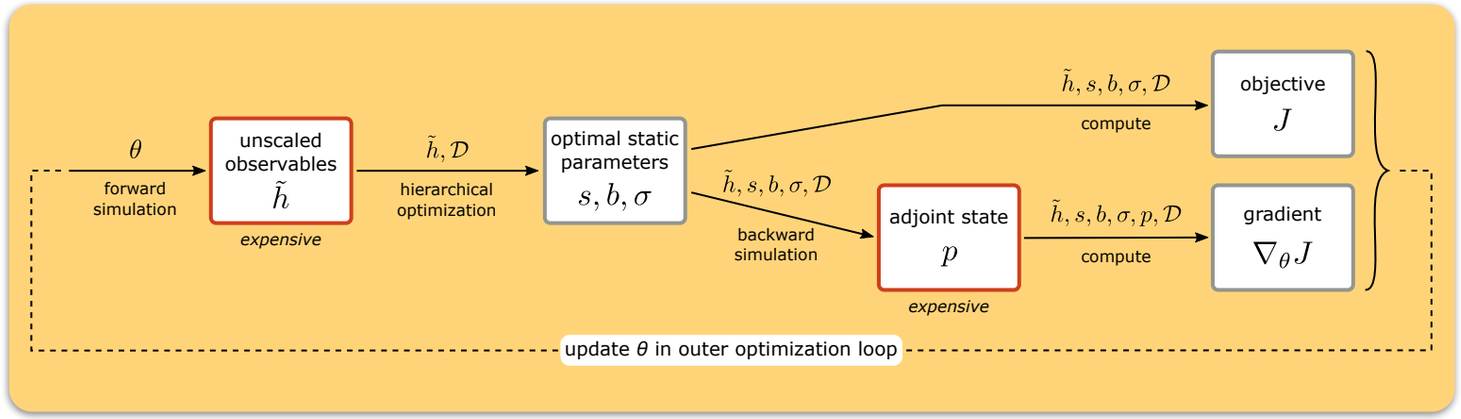
Evaluation (1)

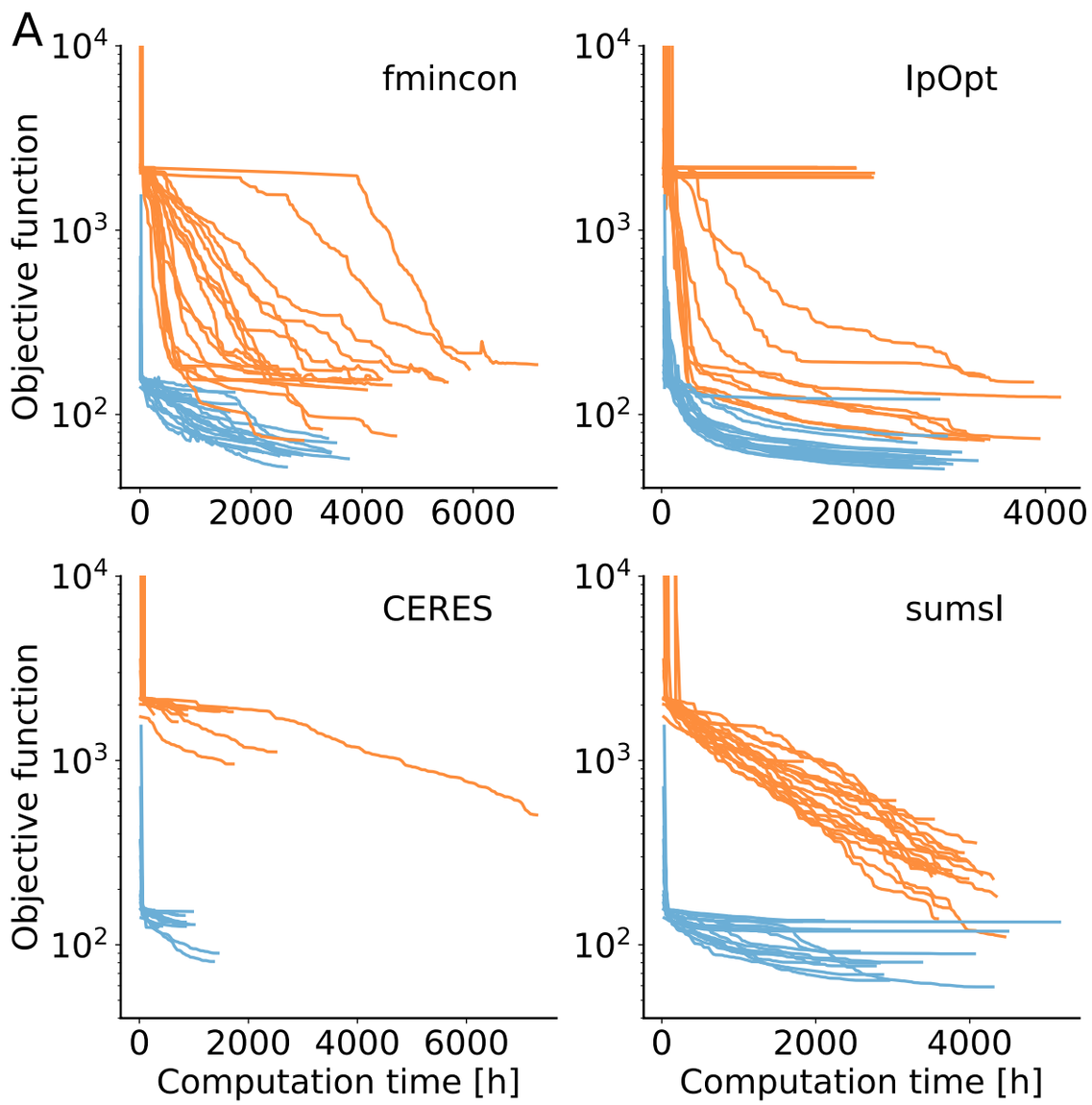


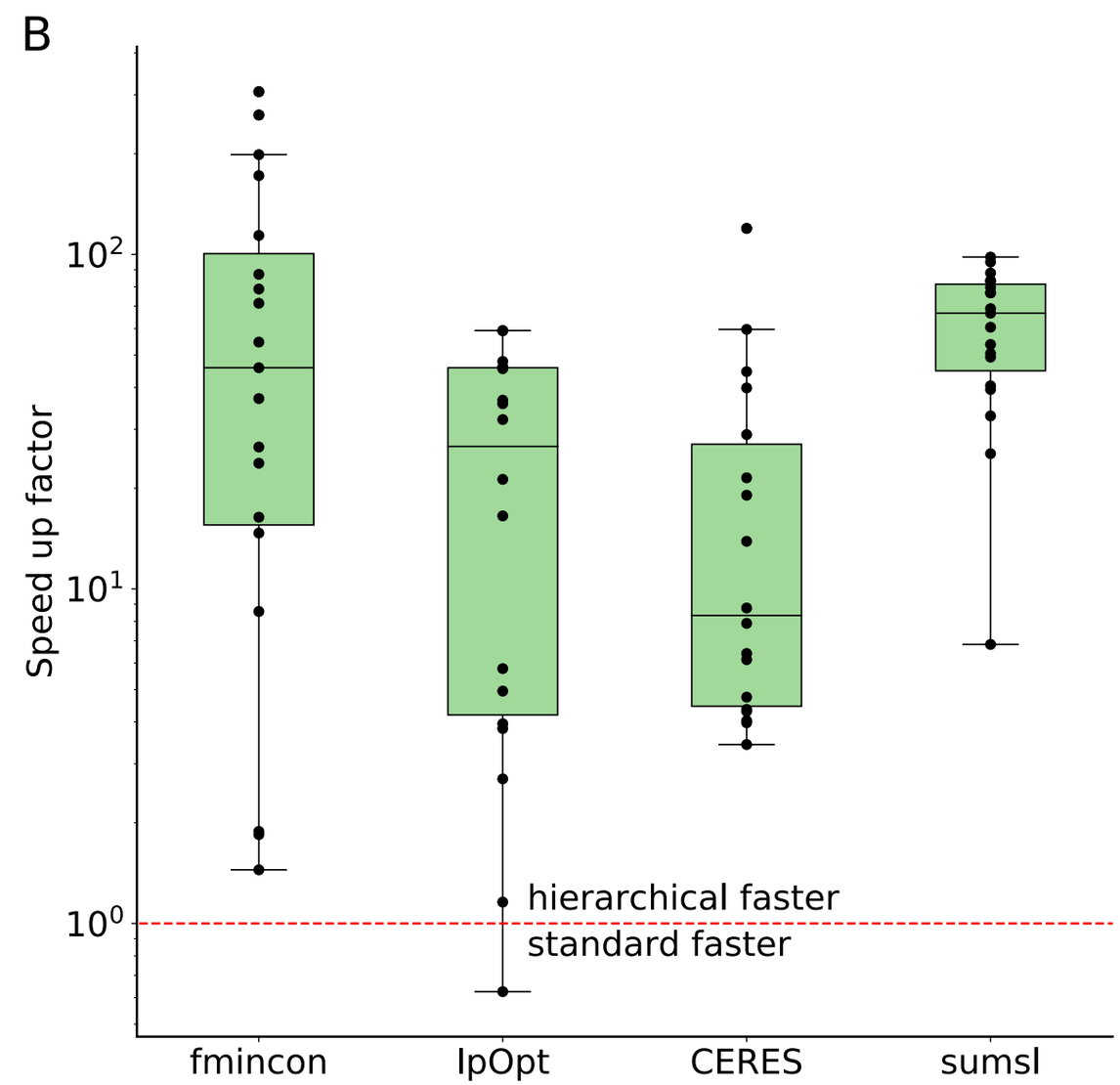
Hierarchical optimisation outperforms standard optimisation methods with respect to convergence and computation time.

Loos et al. Hierarchical optimization for the efficient parametrization of ODE models. Bioinformatics, 34(24):4266-4273, 2018.

Evaluation (2)

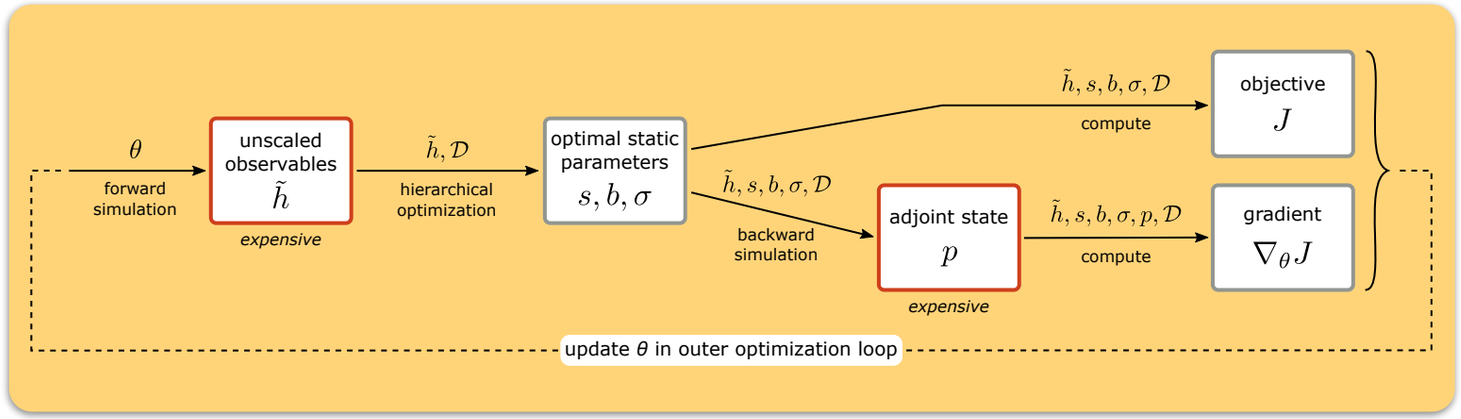


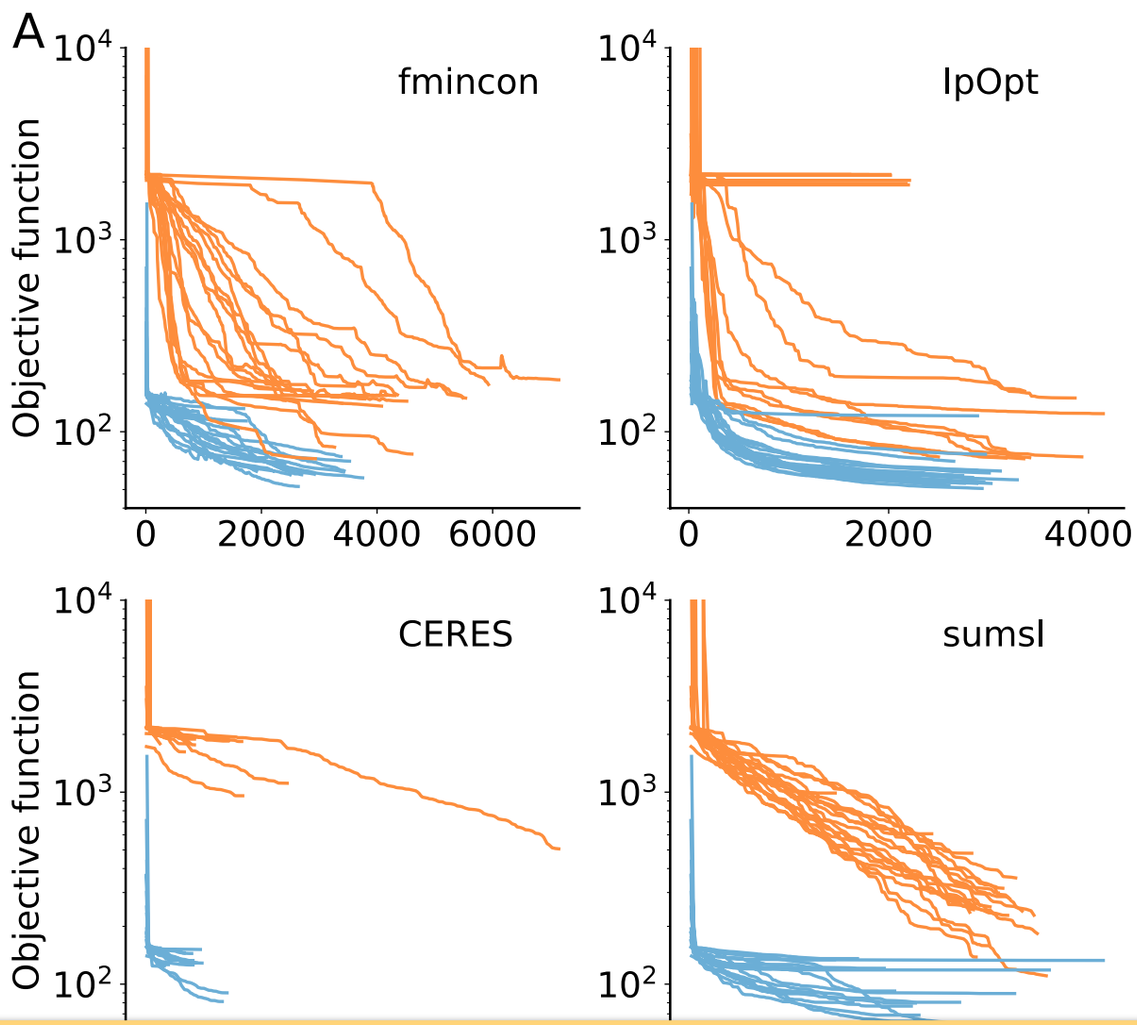


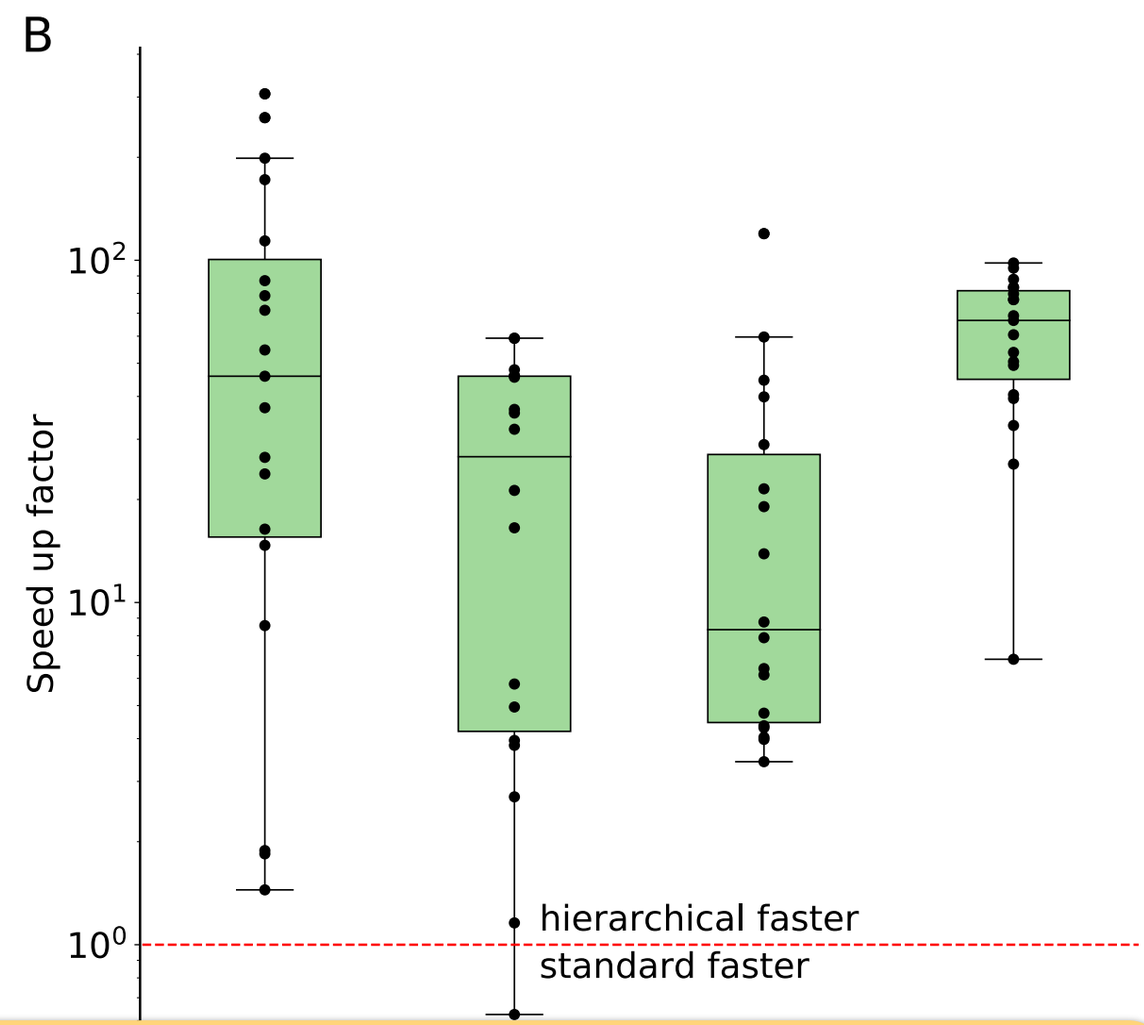


Schmiester et al. Efficient parameterization of large-scale dynamic models based on relative measurements, in preparation.

Evaluation (2)







Hierarchical formulation improves performance of all considered optimisation method.

Schmiester et al. Efficient parameterization of large-scale dynamic models based on relative measurements, in preparation.

Literature

Modelling of relative data

- A. Degasperi, D. Fey, and B. Kholodenko. Performance of objective functions and optimisation procedures for parameter estimation in system biology models. npj Systems Biology and Applications, 3(1), 2017.

Hierarchical optimisation methods

- C. Loos, S. Krause, and J. Hasenauer. Hierarchical optimization for the efficient parametrization of ODE models. Bioinformatics, 34(24):4266-4273, 2018.
- L. Schmiester, Y. Schälte, F. Fröhlich, J. Hasenauer, and D. Weindl. Efficient parameterization of large-scale dynamic models based on relative measurements, *in preparation*