Model-Driven Optimization of Biological Systems

Rafael Arutjunjan

Institute of Physics, University of Freiburg

19th of June 2022

Before We Start

`https://julialang.org/downloads/`

julia Download Documentation Blog Community Learn Research JSoC Sponsor

Download Julia

Star 39,560

Please star us on GitHub. If you use Julia in your research, please cite us. If possible, do consider sponsoring us.

Current stable release: v1.7.3 (May 6, 2022)

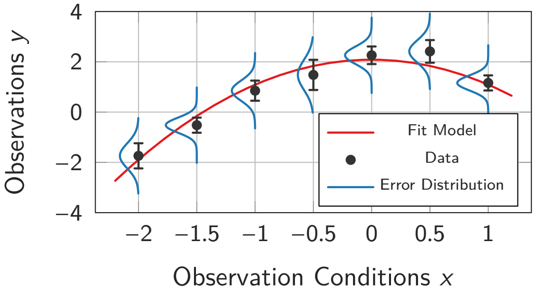
Checksums for this release are available in both MD5 and SHA256 formats.

Windows [help]	64-bit (installer), 64-bit (portable)		32-bit (installer), 32-bit (portable)
macOS x86 (Intel or Rosetta) [help]	64-bit (.dmg), 64-bit (.tar.gz)		
Generic Linux on x86 [help]	64-bit (glibc) (GPG), 64-bit (musl)[1] (GPG)		32-bit (GPG)
Generic Linux on ARM [help]	64-bit (AArch64) (GPG)		32-bit (ARMv7-a hard float) (GPG)
Generic FreeBSD on x86 [help]	64-bit (GPG)		
Source	Tarball (GPG)	Tarball with dependencies (GPG)	GitHub

ODE-Based Models

- A model $y_{\text{model}}(x,\theta)$ is a differentiable function $y_{\text{model}}: \mathscr{X} \times \mathscr{M} \longrightarrow \mathscr{Y}$
- $\mathscr{X}$ and $\mathscr{Y}$ respectively denote the space of independent and dependent variables, $\mathscr{M}$ denotes the parameter space
- Models in life sciences often formulated in terms of "chemical" reactions with different rates
- Write down network graph, then convert into system of Ordinary Differential Equations (ODEs)
- Partial Differential Equations or Stochastic Differential Equations also possible, same idea

What is a dataset?



- Dataset consists of observations $y_i \in \mathscr{Y}$, observation conditions $x_i \in \mathscr{X}$ and a specification of the uncertainty in the data points
- Consider the collection of all N observations as a single point $y_{\text{data}} := (y_1, \dots, y_N) \in \mathscr{Y}^N$ in the “data space”

Maximum Likelihood Estimation

How do we decide which parameters $\theta \in \mathcal{M}$ best describe the data?

- Use Gaussian likelihood L defined by

$$L(\mathrm{data}\,|\,\theta) := \sqrt{\frac{\det\left(\Sigma^{-1}\right)}{(2\pi)^N}}\,\exp\left\{-\frac{1}{2}\left(y_{\mathrm{data}} - h(\theta)\right)^\top \Sigma^{-1}\left(y_{\mathrm{data}} - h(\theta)\right)\right\}$$

$$\text{where} \qquad h(\theta) := \left(y_{\mathrm{model}}(x_1, \theta), \ldots, y_{\mathrm{model}}(x_N, \theta)\right) \in \mathcal{Y}^N$$

- Maximum likelihood estimate (MLE) defined by

$$\nabla_\theta L = 0 \qquad \text{i.e.} \qquad \theta_{\mathrm{MLE}} := \operatorname*{arg\,max}_{\theta \in \mathcal{M}} L(\theta)$$

And Now for Something Completely Different

github.com/RafaelArutjunjan/LivingMaterials2022

```
julia> using Pkg
julia> Pkg.add("Pluto")
julia> using Pluto
julia> Pluto.run()
```

Parameter Uncertainty

How sure are we that the parameters θ_{MLE} we found are correct?

- For **linearly parametrized** models, confidence regions are always perfect ellipsoids

$\Longrightarrow$ Parameter covariance can be described using positive-definite matrix

- For **non-linearly parametrized** models, confidence regions exhibit complicated, distorted shapes

$\Longrightarrow$ Shape cannot be encoded in a matrix

The Fisher Information Matrix g

- Computes expected curvature (i.e. Hessian) of the log-likelihood $\ln(L)$

$$g_{ab}(\theta) := -\mathbb{E}\left(\frac{\partial^2 \ln(L)}{\partial\theta^a \partial\theta^b}\right)$$

- For Gaussian likelihoods, it can be conveniently computed via [3, 16]

$$g(\theta) = \left(\frac{\partial h}{\partial\theta}\right)^\top \Sigma^{-1} \left(\frac{\partial h}{\partial\theta}\right)$$

- Parameter covariance matrix is given by the inverse Fisher information

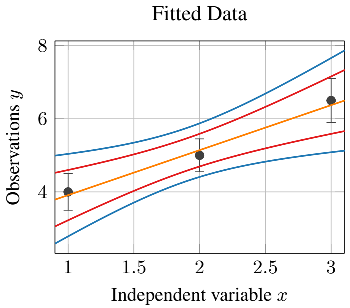
$$\Sigma_{\mathscr{M}} \approx g^{-1}(\theta_{\mathrm{MLE}})$$

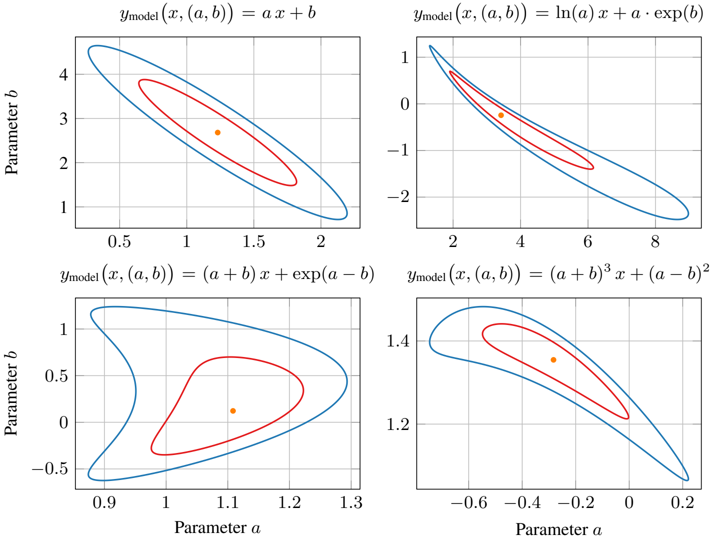
- In general, $\Sigma_{\mathscr{M}}$ is a function of position in the parameter space!

Confidence Regions for Linear vs Non-Linear Models

Toy data

x_i	y_i	σ_i
1	4	0.5
2	5	0.45
3	6.5	0.6

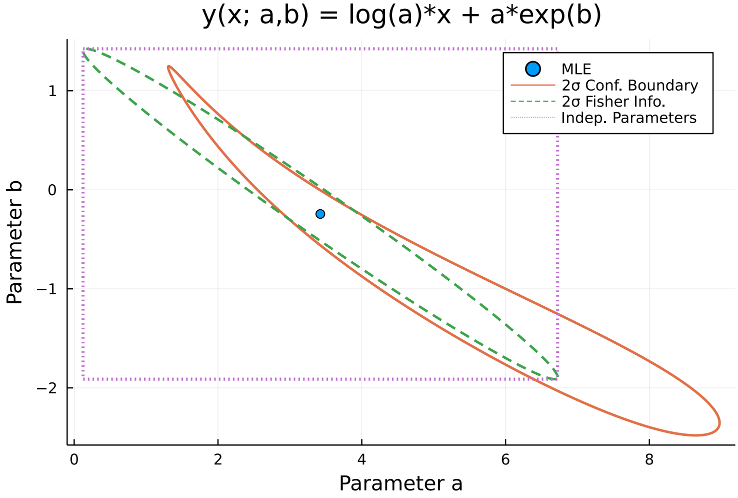




Common (But BAD) Practice

1. Ignore non-linearity in model and just focus on Fisher information matrix
2. Pretend that parameters don't affect each other, i.e. Fisher matrix diagonal
3. Take square root of diagonal entries of inverse matrix as symmetric parameter uncertainties
4. Publish

Exact Confidence Regions vs Approximations



Typically published as

$$a_{\mathrm{MLE}} = 3.4 \pm 1.3$$

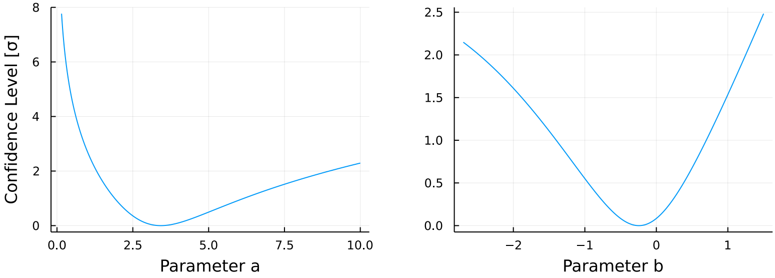
$$b_{\mathrm{MLE}} = -0.24 \pm 0.67$$

Guess what – ODE models are always* non-linear with respect to their parameters!

* with the exception of very few carefully constructed examples

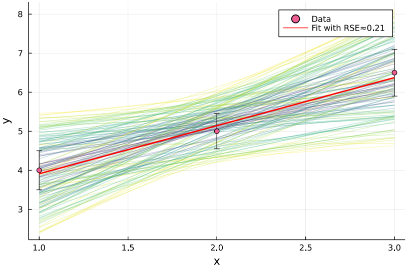
Profile Likelihood

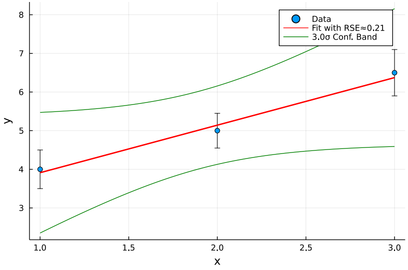
- Proper one-dimensional assessment of uncertainty in the parameters
- Fix a single parameter at a specific value and re-optimize the remaining parameters, then compute likelihood [12, 14]
- Walks along a “trajectory of slowest decent” in parameter space



Confidence Bands

- Why care about parameter uncertainty in the first place?
- Required to estimate uncertainty in model prediction
- Prediction uncertainty is computed by evaluating model for parameter configurations in confidence region





Can you always determine the model parameters from data?

Nope.

Why not?

- **Structural non-identifiability** "model is badly defined":
  $y_{\text{model}}\big(x, (a, b)\big) = a \cdot b \cdot x$

$\implies$ $\exists$ direction in parameter space along which $L(\theta)$ does not change

- Rothenberg [13]: Model identifiable at θ $\iff$ $\det\big(g(\theta)\big) \neq 0$
- **Practical non-identifiability** "insufficient / bad data":

$$\mathrm{A} \xleftarrow{k_A} \mathrm{E} \xrightarrow{k_B} \mathrm{B}$$

$\implies$ Cannot determine k_A if only B is measured, results in non-compact confidence regions

SIR example

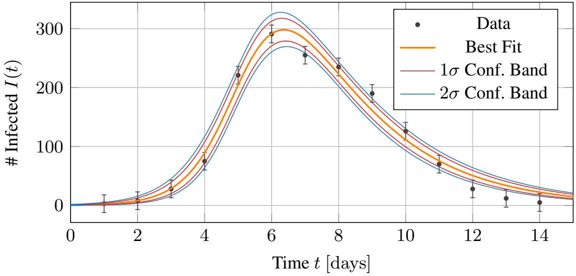
- Model given by: $S \xrightarrow{\beta} I \xrightarrow{\gamma} R$

$$\implies \frac{\mathrm{d}S}{\mathrm{d}t} = -\beta S(t) I(t), \quad \frac{\mathrm{d}I}{\mathrm{d}t} = \beta S(t) I(t) - \gamma I(t), \quad \frac{\mathrm{d}R}{\mathrm{d}t} = \gamma I(t)$$

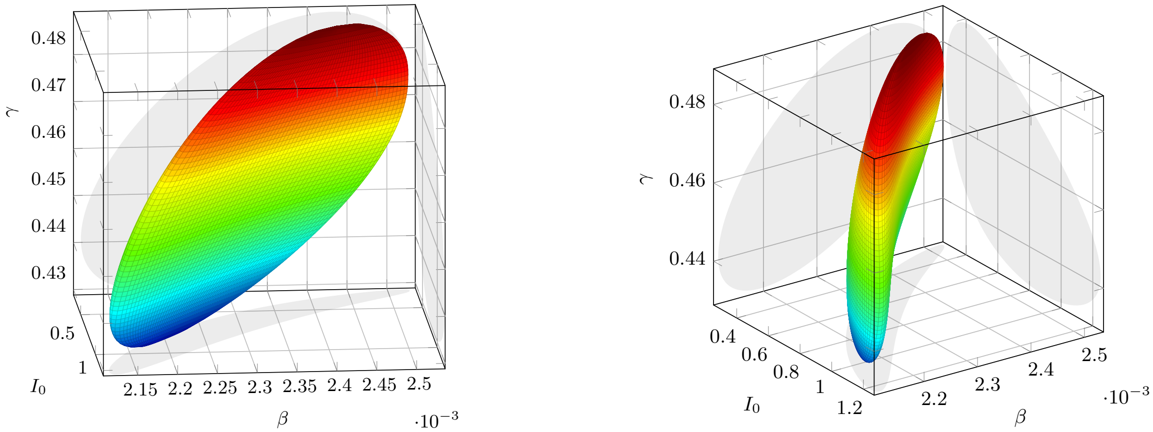
- Infection dataset from English Boarding School in 1978 [2, 11]

time [days]	1	2	3	4	5	6	7	8	9	10	11	12	13	14
# Infected	3	8	28	75	221	291	255	235	190	126	70	28	12	5

Total number of pupils at school: 763
Standard deviation assumed as $\sigma = 15$



SIR Confidence Regions



- Non-linearity in parameters visible from distorted confidence region

Further Questions?

✉ rafael.arutjunjan@fdm.uni-freiburg.de

/RafaelArutjunjan/LivingMaterials2022

/RafaelArutjunjan/InformationGeometry.jl

References I

[1] Amari, S. ; Nagaoka, H. : *Methods of Information Geometry*. American Mathematical Society (Translations of mathematical monographs). https://books.google.de/books?id=vc2FWSo7wLUC. – ISBN 9780821843024

[2] Anonymous: Influenza in a boarding school. In: *British Medial Journal* (1978), März, 587. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1603269/pdf/brmedj00115-0064.pdf

[3] Arutjunjan, R. : *On the Geometric Foundation of Parameter Inference*, Friedrich-Alexander University Erlangen-Nürnberg, Master's thesis, August 2020. https://github.com/RafaelArutjunjan/Master-Thesis

[4] Arutjunjan, R. : RafaelArutjunjan/InformationGeometry.jl: v1.11.3. (2022), März. http://dx.doi.org/10.5281/zenodo.6358843. – DOI 10.5281/zenodo.6358843

[5] Dandekar, R. ; Rackauckas, C. ; Barbastathis, G. : A Machine Learning-Aided Global Diagnostic and Comparative Tool to Assess Effect of Quarantine Control in COVID-19 Spread. In: *Patterns* 1 (2020), Nr. 9, 100145. http://dx.doi.org/https://doi.org/10.1016/j.patter.2020.100145. – DOI https://doi.org/10.1016/j.patter.2020.100145. – ISSN 2666–3899

[6] Kramer, B. P. ; Viretta, A. U. ; Baba, M. D.-E. ; Aubel, D. ; Weber, W. ; Fussenegger, M. : An engineered epigenetic transgene switch in mammalian cells. In: *Nature Biotechnology* 22 (2004), Jul, Nr. 7, 867-870. http://dx.doi.org/10.1038/nbt980. – DOI 10.1038/nbt980. – ISSN 1546–1696

[7] Mogensen, P. K. ; Riseth, A. N.: Optim: A mathematical optimization package for Julia. In: *Journal of Open Source Software* 3 (2018), Nr. 24, S. 615. http://dx.doi.org/10.21105/joss.00615. – DOI 10.21105/joss.00615

[8] Müller, K. ; Engesser, R. ; Metzger, S. ; Schulz, S. ; Kämpf, M. M. ; Busacker, M. ; Steinberg, T. ; Tomakidi, P. ; Ehrbar, M. ; Nagy, F. ; Timmer, J. ; Zubriggen, M. D. ; Weber, W. : A red/far-red light-responsive bi-stable toggle switch to control gene expression in mammalian cells. In: *Nucleic Acids Research* 41 (2013), 01, Nr. 7, e77-e77. http://dx.doi.org/10.1093/nar/gkt002. – DOI 10.1093/nar/gkt002. – ISSN 0305–1048

[9] Onzon, E. : Multivariate Cramér–Rao inequality for prediction and efficient predictors. In: *Statistics & Probability Letters* 81 (2011), Nr. 3, 429 - 437. http://dx.doi.org/https://doi.org/10.1016/j.spl.2010.12.007. – DOI https://doi.org/10.1016/j.spl.2010.12.007. – ISSN 0167–7152

[10] Rackauckas, C. ; Nie, Q. : DifferentialEquations.jl – A Performant and Feature-Rich Ecosystem for Solving Differential Equations in Julia. In: *Journal of Open Source Software* (2017). http://dx.doi.org/http://doi.org/10.5334/jors.151. – DOI http://doi.org/10.5334/jors.151

References II

[11] RAISSI, M. ; RAMEZANI, N. ; SESHAIYER, P. : On parameter estimation approaches for predicting disease transmission through optimization, deep learning and statistical inference methods. In: *Letters in Biomathematics* 6 (2019), Jan., Nr. 2, 1–26. http://dx.doi.org/10.1080/23737867.2019.1676172. – DOI 10.1080/23737867.2019.1676172. – ISSN 2373–7867

[12] RAUE, A. ; KREUTZ, C. ; MAIWALD, T. ; BACHMANN, J. ; SCHILLING, M. ; KLINGMÜLLER, U. ; TIMMER, J. : Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. In: *Bioinformatics* 25 (2009), 06, Nr. 15, 1923-1929. http://dx.doi.org/10.1093/bioinformatics/btp358. – DOI 10.1093/bioinformatics/btp358. – ISSN 1367–4803

[13] ROTHENBERG, T. J.: Identification in Parametric Models. In: *Econometrica* 39 (1971), Nr. 3, 577–591. http://www.jstor.org/stable/1913267. – ISSN 00129682, 14680262

[14] SEBER, G. ; WILD, C. : *Nonlinear Regression*. Wiley (Wiley Series in Probability and Statistics). https://books.google.de/books?id=YBYlCpBNo_cC. – ISBN 9780471471356

[15] TRANSTRUM, M. K. ; MACHTA, B. ; BROWN, K. ; DANIELS, B. C. ; MYERS, C. R. ; SETHNA, J. P.: Sloppiness and Emergent Theories in Physics, Biology, and Beyond. (2015). https://arxiv.org/pdf/1501.07668.pdf

[16] TRANSTRUM, M. K. ; MACHTA, B. B. ; SETHNA, J. P.: Geometry of nonlinear least squares with applications to sloppy models and optimization. 83 (2011), March, Nr. 3, S. 036701. http://dx.doi.org/10.1103/PhysRevE.83.036701. – DOI 10.1103/PhysRevE.83.036701

[17] WILKS, S. S.: The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. In: *Ann. Math. Statist.* 9 (1938), 03, Nr. 1, 60–62. http://dx.doi.org/10.1214/aoms/1177732360. – DOI 10.1214/aoms/1177732360

https://knowyourmeme.com/memes/this-is-fine

Precise Definition of Confidence Regions

- A theorem due to Wilks [17] states that in the large sample limit $(N \longrightarrow \infty)$, the likelihood ratio is distributed according to

$$2\Big(\log\big(L(\theta_{\mathrm{MLE}})\big) - \log\big(L(\theta)\big)\Big) \sim \chi^2_{\dim \mathscr{M}}.$$

- Therefore, the inverse cdf of the χ^2 distribution with $\dim \mathscr{M}$ degrees of freedom can be used to set a threshold above which configurations $\theta \in \mathscr{M}$ are rejected at a confidence level $q \in (0,1)$

$$\mathscr{C}_q := \left\{ \theta \in \mathscr{M} \,\middle|\, 2\Big(\log\big(L(\theta_{\mathrm{MLE}})\big) - \log\big(L(\theta)\big)\Big) \leq \mathrm{icdf}\big(\chi^2_{\dim \mathscr{M}}, q\big) \right\}$$

Definition of the Profile Likelihood

- Profile Likelihood for a (subset of) parameter(s) ξ

$$\mathrm{PL}_\delta(\xi) := \sup_\delta L(\mathrm{data} \,|\, \xi, \delta)$$

for $\theta = (\xi, \delta)$ with δ some nuisance parameters.

- Corresponds to a projection of the full parameter space $\mathscr{M}$ onto a subspace
- Profile-based confidence intervals defined as

$$\mathrm{CI}_q := \left\{ \xi \,\middle|\, 2\Big(\log\big(L(\theta_{\mathsf{MLE}})\big) - \log\big(\mathrm{PL}_\delta(\xi)\big)\Big) \leq \mathrm{icdf}\big(\chi^2_{\dim \mathscr{M}}, q\big) \right\}$$

Correspondence Between Confidence Regions and Profiles