Cosmology with Bayesian statistics and information theory

Lecture 2: Probabilistic computations

… a.k.a. *how much do I know about the likelihood?*

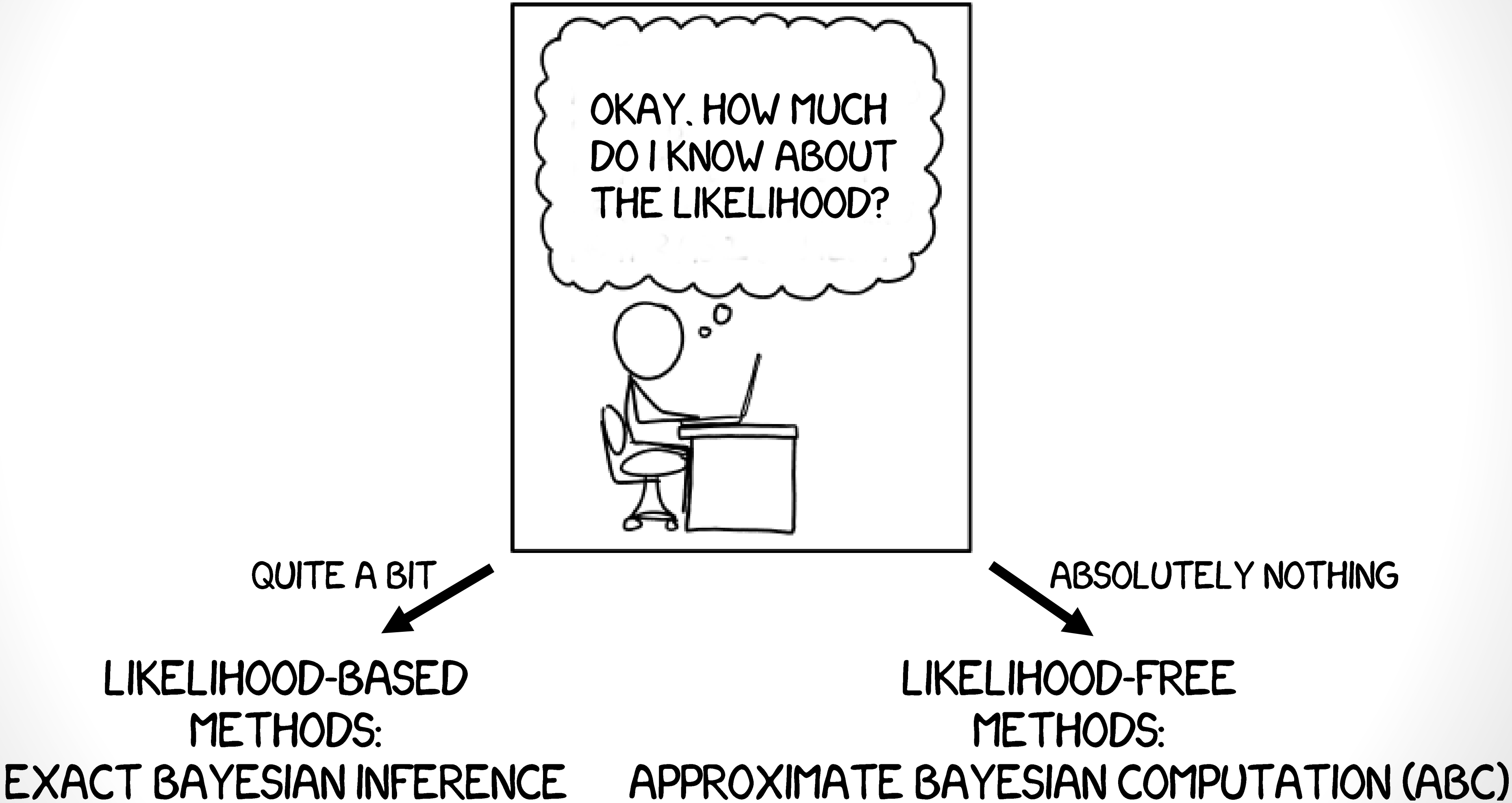
Florent Leclercq

Institute of Cosmology and Gravitation, University of Portsmouth

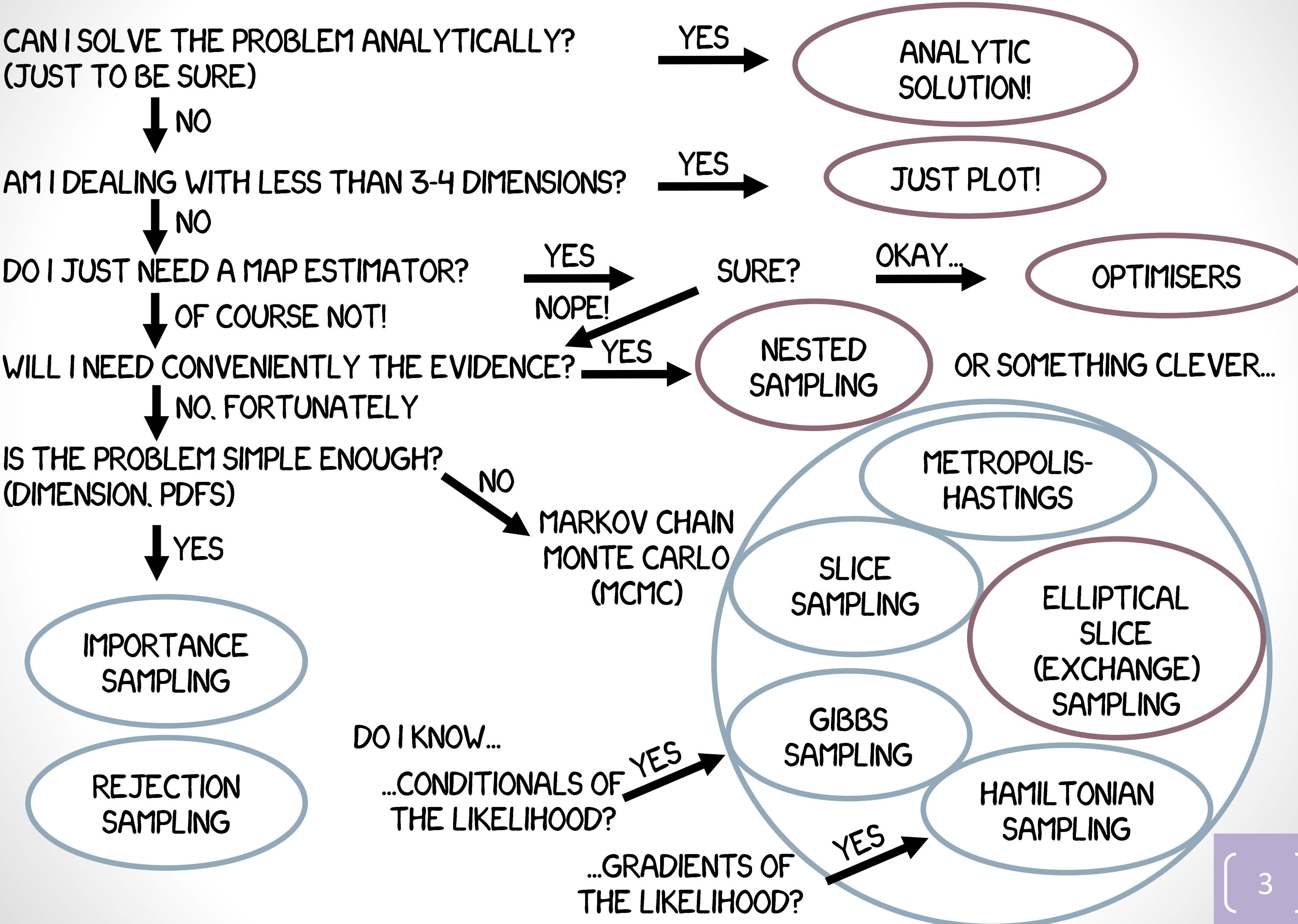
http://icg.port.ac.uk/~leclercq/

March 8th, 2017

Probabilistic computations: two approaches
(a very personal view)



LIKELIHOOD-BASED METHODS: EXACT BAYESIAN INFERENCE



LIKELIHOOD-FREE METHODS: APPROXIMATE BAYESIAN COMPUTATION (ABC)

DO I REALLY KNOW NOTHING ABOUT THE LIKELIHOOD?

→ NO, I'M JUST LAZY → ABC IS NOT CARTE BLANCHE REPLACEMENT FOR LIKELIHOOD-BASED METHODS. I GO BACK TO THE PREVIOUS SLIDE.

↓ YES

DO I KNOW A SUFFICIENT SUMMARY STATISTIC OF MY DATA?

→ YES → GREAT, I'M USING IT IN THE FOLLOWING!

↓ NO

DO I HAVE A REASONABLE CHANCE TO HIT THE DATA "BY CHANCE" WITH MY GENERATIVE MODEL? (STRONG PRIOR, LOW DIMENSION, HIGH TOLERANCE)

→ YES

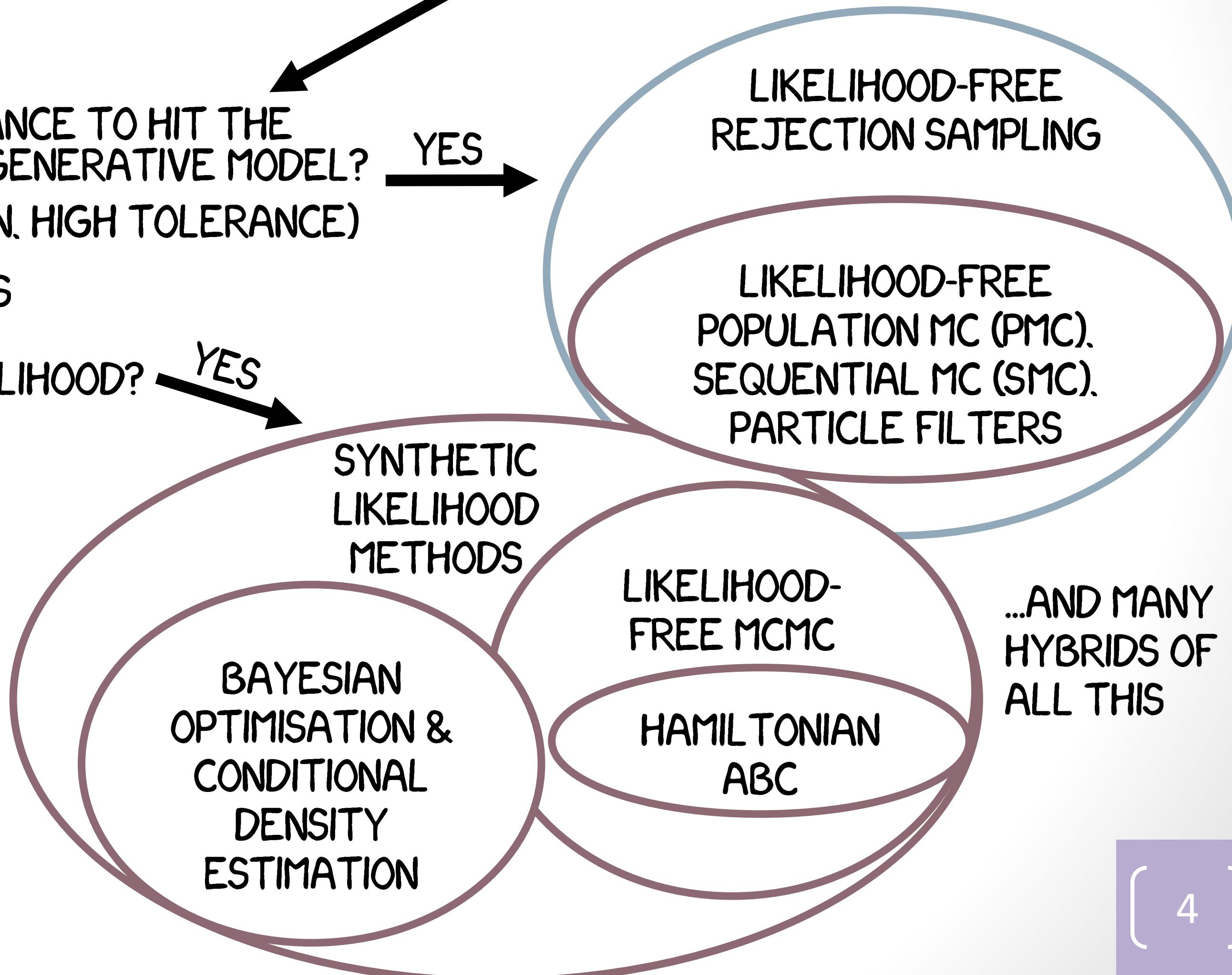
↓ NO, IT'S HOPELESS

CAN I BUILD A SYNTHETIC LIKELIHOOD?

→ YES

↓ NO

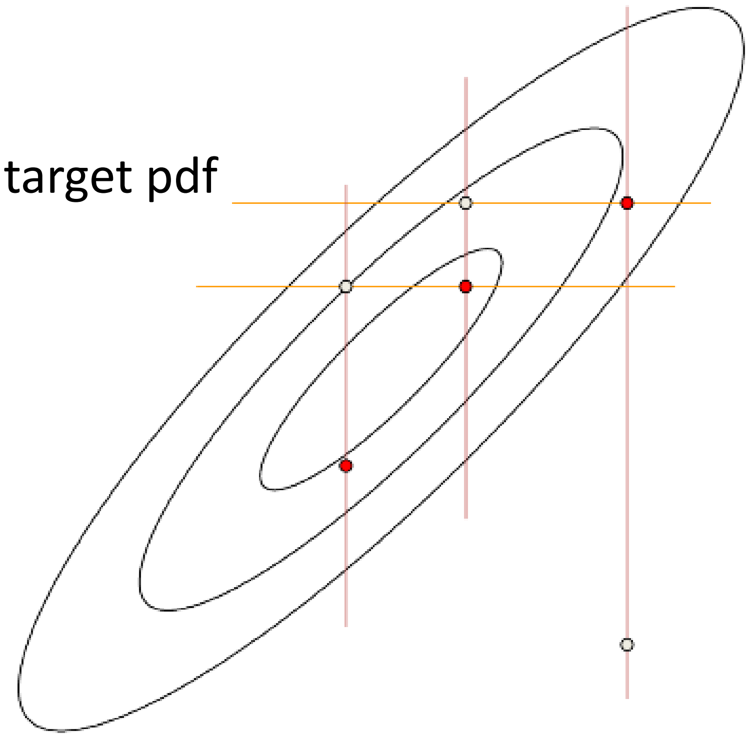
HOPELESS PROBLEM. I TAKE A BREAK, THEN THINK MORE.



MCMC beyond Metropolis-Hastings

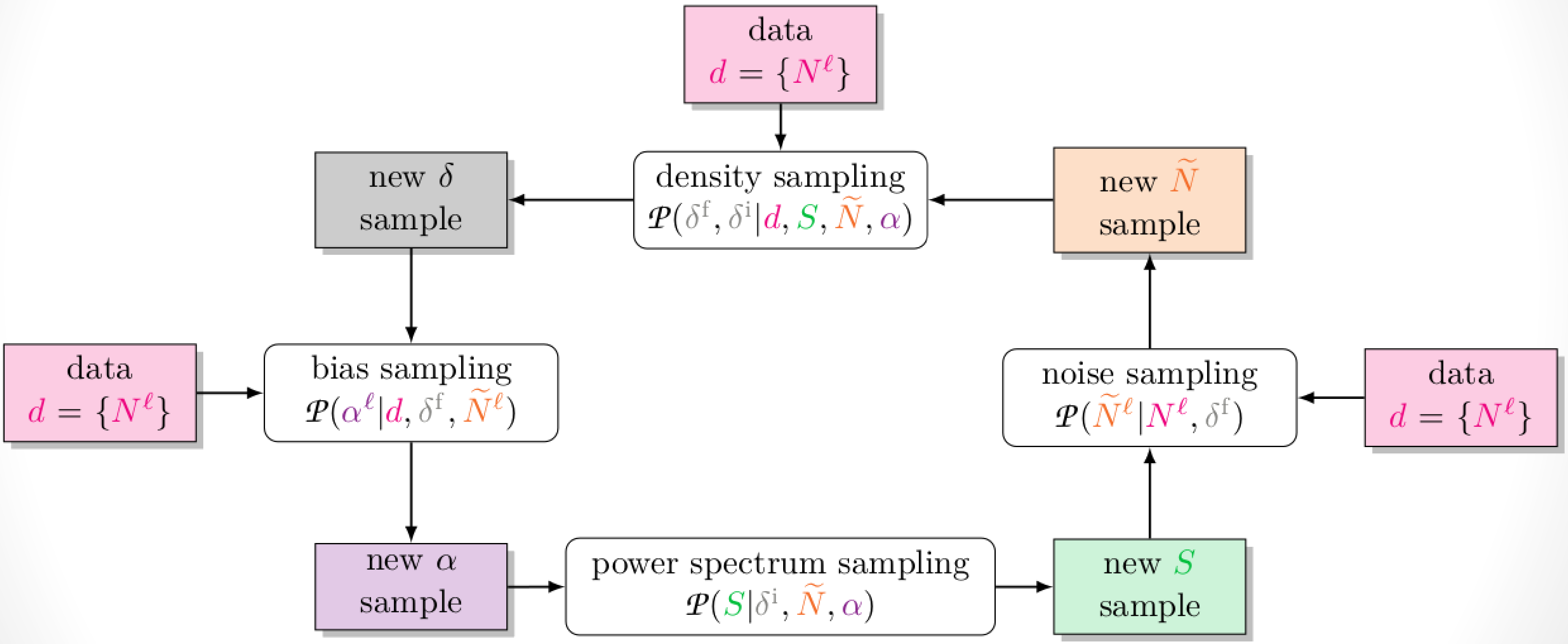
- Shortcomings of standard Metropolis-Hastings:
  - Tuning of proposal distributions
  - Curse of dimensionality

- Gibbs sampling:
  - Uses conditionals of the target pdf



Modular probabilistic programming: example

ARES: *Algorithm for REconstruction and Sampling*

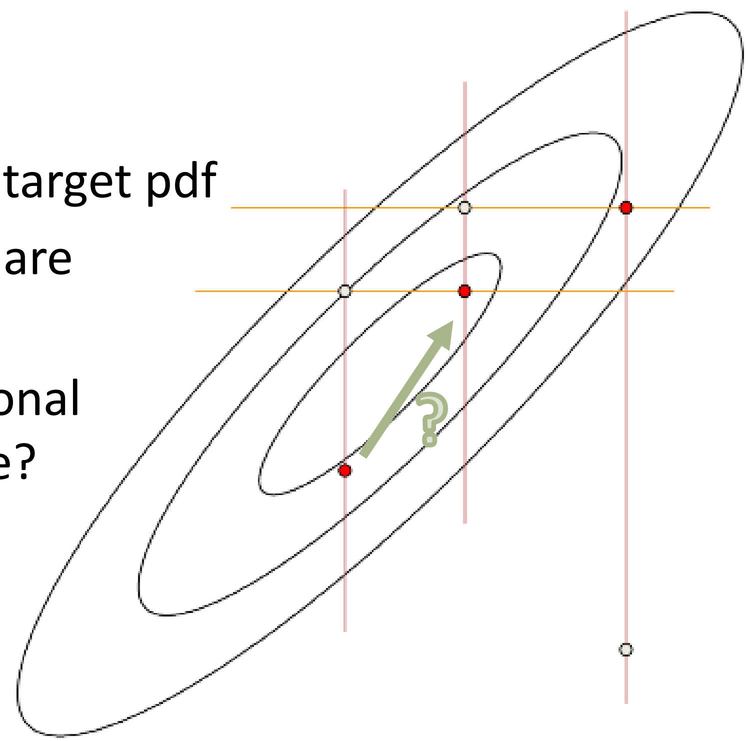


Jasche, Kitaura, Wandelt & Enßlin 2010, arXiv:0911.2493
Jasche & Wandelt 2013, arXiv:1306.1821

MCMC beyond Metropolis-Hastings

- Shortcomings of standard Metropolis-Hastings:
  - Tuning of proposal distributions
  - Curse of dimensionality

- Gibbs sampling:
  - Uses conditionals of the target pdf
  - Inefficient if parameters are strongly correlated
  - How does one take diagonal steps in parameter space?



Hamiltonian (Hybrid) Monte Carlo

- Use classical mechanics to solve statistical problems!
  - The potential: $\psi(\mathbf{x}) \equiv -\ln p(\mathbf{x})$
  - The Hamiltonian: $H(\mathbf{x},\mathbf{p}) \equiv \frac{1}{2}\mathbf{p}^\intercal \mathbf{M}^{-1}\mathbf{p} + \psi(\mathbf{x})$

$$(\mathbf{x},\mathbf{p}) \Rightarrow \left\{ \begin{aligned} \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} &= \frac{\partial H}{\partial \mathbf{p}} = \mathbf{M}^{-1}\mathbf{p} \\ \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} &= -\frac{\partial H}{\partial \mathbf{x}} = -\frac{\mathrm{d}\psi(\mathbf{x})}{\mathrm{d}\mathbf{x}} \end{aligned} \right\} \Rightarrow (\mathbf{x}',\mathbf{p}')$$

gradients of the pdf

$$a(\mathbf{x}',\mathbf{x}) = \mathrm{e}^{-(H'-H)} = 1$$

acceptance ratio unity

- HMC **beats the curse of dimensionality** by:
  - Exploiting gradients
  - Using conservation of the Hamiltonian

Duane *et al.* 1987, Phys. Lett. B **195**, 2