Static computational budget optimization with stochastic simulations

Jean-François Paquet July 8, 2024 **PAPER • OPEN ACCESS**

[arXiv:2301.08385]

Computational budget optimization for Bayesian parameter estimation in heavy-ion collisions

Brandon Weiss¹ D, Jean-François Paquet^{1,2} D and Steffen A Bass¹ Published 16 May 2023 • © 2023 The Author(s). Published by IOP Publishing Ltd Journal of Physics G: Nuclear and Particle Physics, Volume 50, Number 6 Citation Brandon Weiss *et al* 2023 *J. Phys. G: Nucl. Part. Phys.* 50 065104

DOI 10.1088/1361-6471/acd0c7



"Inverse Problems and Uncertainty Quantification in Nuclear Physics" Workshop



Ultrarelativistic heavy-ion collisions



Ref.: ALICE, CERN



- Energy-momentum tensor of plasma: $T^{\mu\nu} = \epsilon u^{\mu}u^{\nu} (P(\epsilon) + \Pi)(g^{\mu\nu} u^{\mu}u^{\nu}) + \pi^{\mu\nu}$
- Conservation of energy and momentum: $\partial_{\nu}T^{\mu\nu} = 0$
- Transient relativistic viscous hydrodynamics

 $\tau_{\pi}\Delta^{\mu\nu}_{\alpha\beta}\dot{\pi}^{\alpha\beta} + \pi^{\mu\nu} = 2 \, \eta(T)(\partial^{\mu}u^{\nu} + \cdots) + (2^{nd} \text{ order}); \quad \tau_{\Pi}\dot{\Pi} + \Pi = -\zeta(T) \, \partial_{\mu} \, u^{\mu} + (2^{nd} \text{ order});$



- Energy-momentum tensor of plasma: $T^{\mu\nu} = \epsilon u^{\mu}u^{\nu} (P(\epsilon) + \Pi)(g^{\mu\nu} u^{\mu}u^{\nu}) + \pi^{\mu\nu}$
- Conservation of energy and momentum: $\partial_{\nu}T^{\mu\nu} = 0$
- Transient relativistic viscous hydrodynamics

 $\tau_{\pi}\Delta^{\mu\nu}_{\alpha\beta}\dot{\pi}^{\alpha\beta} + \pi^{\mu\nu} = 2 \,\eta(T)(\partial^{\mu}u^{\nu} + \cdots) + (2^{nd} \text{ order}); \quad \tau_{\Pi}\dot{\Pi} + \Pi = -\zeta(T) \,\partial_{\mu} \,u^{\mu} + (2^{nd} \text{ order});$





- Functional description of the physics of heavy-ion collisions using: lattice QCD, relativistic hydrodynamics&transport, perturbative QCD, ...
- Model parameters: unknown or uncertain quantities
 - Shear&bulk viscosity of the plasma
 - Perhaps equation of state
 - Many parameters in the "initial stage" and in transition between stages



Data available for handful of nuclei and \sqrt{s} (e.g. Au-Au $\sqrt{s_{NN}}$ =0.2 TeV, Pb-Pb $\sqrt{s_{NN}}$ = 2.76 TeV)

- Large number of measured observables
 - Many energy/momentum bins
 - Many centrality bins

Percent-level precision is common

Model-data comparison

Constraints from Bayesian inference:

posterior(*param*) $\propto exp\left(-\frac{1}{2}\sum_{observables}\frac{(model(param) - data)^2}{(uncertainty)^2}\right)$ Constraints on Challenge: many parameters (~10-20)

Constraints on parameters

Model prediction for given set of model parameters









- For one choice of model parameters:
 - Simulate 10³-10⁴ collisions (expt: 10⁶) [model is stochastic]
 - Simulation: a few core-minute/collision
 - = ~100-1000 core-hours per param
- 1k param samples $\rightarrow \sim 10^5 \cdot 10^6$ core-hour
- 10k param samples $\rightarrow \sim 10^6 \cdot 10^7$ core-hour

Model-data comparison

MODEL TRAINING & OE

A 20-dimensional model parameter space with 1,000 training points

Au+Au	Hydro events per design	Avg. hadronic events per hydro
200 GeV	1,000	1,000
19.6 GeV	2,000	4,000
7.7 GeV	2,000	8,000

Cpen Science Grid delivered 5 million CPU hours for the data generation

Chun Shen (Wayne State)

Inverse Problems and Uncertair

- For one choice of model parameters:
 - Simulate 10³-10⁴ collisions (expt: 10⁶) [model is stochastic]
 - Simulation: a few core-minute/collision
 - = ~100-1000 core-hours per param

- 1k param samples $\rightarrow \sim 10^5 \cdot 10^6$ core-hour
- 10k param samples $\rightarrow \sim 10^6 \cdot 10^7$ core-hour

Chun's talk this morning

Emulation

posterior(\overline{param}) $\propto prior(\overline{param}) \times exp\left(-\frac{1}{2}\left(\overline{Model}(\overline{param}) - \overline{D}\right)^T \operatorname{Covar}^{-1}\left(\overline{Model}(\overline{param}) - \overline{D}\right)\right)$

- Posterior is high-dimensional: expensive to sample
- Solution (that we use): replace model by emulators (Gaussian processes)
- Emulator covariance kernel:

$$k(\mathbf{x}_p, \mathbf{x}_q) = k_{\exp}(\mathbf{x}_p, \mathbf{x}_q) + k_{\operatorname{noise}}(\mathbf{x}_p, \mathbf{x}_q)$$

"Interpolation component":

$$k_{\exp}(\mathbf{x}_p, \mathbf{x}_q) = C^2 \exp\left(-\frac{1}{2} \sum_{i=1}^{s} \frac{|x_{p,i} - x_{q,i}|^2}{l_i^2}\right)$$

"Statistical component":

$$k_{\text{noise}}(\mathbf{x}_p, \mathbf{x}_q) = \sigma_{\text{noise}}^2 \delta_{p,q}$$



Emulation with stochastic simulations

Fig. ref.: https://scikitlearn.org/0.17/auto examples/gaussian process/plot gp regression.html



Larger stat. uncertainty

- Given computational budget:
 - *M_{event}* = collisions per parameter sample
 - N_{param samples} = number of parameter samples
 - Budget = $M_{event} \times N_{param \ samples}$



- Given computational budget:
 - *M_{event}* = collisions per parameter sample
 - N_{param samples} = number of parameter samples
 - Budget = $M_{event} \times N_{param \ samples}$
- For given budget, what is the optimal M_{event} and $N_{param \ samples}$?
 - Optimal = minimizes uncertainty on parameters
- "Rule of thumb"?:

 $N_{param \ samples} \sim 10 \times (number \ of \ model \ parameters)$







 $M_{event} \times N_{param \, samples}$ (budget) fixed





Trade-off in emulation with closure tests







Observable: transverse anisotropy

Based on figures by Derek Teaney, CMS Event display of $\frac{dE}{d\phi}$ in calorimeter Spatial anisotropy Momentum anisotropy Energy deposition **Hydrodynamics** Hadronic transport **Early dynamics**

Transverse initial energy density: $\varepsilon_n \mathrm{e}^{\mathrm{i}n\Phi_n} = \frac{\int_0^\infty \mathrm{d}rr \int_0^{2\pi} \mathrm{d}\phi \ r^n \epsilon(r, \phi) \mathrm{e}^{\mathrm{i}n\phi}}{\int_0^\infty \mathrm{d}rr \int_0^{2\pi} \mathrm{d}\phi \ r^n \epsilon(r, \phi)}$ $\langle \varepsilon_n \rangle = \frac{1}{M_{\rm ev}} \sum_{i=1}^{M_{\rm ev}} \varepsilon_n \{ \text{event } j \}$

Transverse momentum distribution of hadrons:

$$\frac{dN}{d\phi} = \frac{N}{2\pi} \left[1 + 2\sum_{n} v_n \cos(n(\phi - \Phi_n)) \right]$$



Observable: transverse anisotropy



INITIAL ENERGY DEPOSITION (TRENTO)

Parameterization for energy deposition at $\tau = 0^+$





p = -1





$$\varepsilon_n \mathrm{e}^{\mathrm{i}n\Phi_n} = \frac{\int_0^\infty \mathrm{d}rr \int_0^{2\pi} \mathrm{d}\phi \ r^n \epsilon(r, \phi) \mathrm{e}^{\mathrm{i}n\phi}}{\int_0^\infty \mathrm{d}rr \int_0^{2\pi} \mathrm{d}\phi \ r^n \epsilon(r, \phi)}$$
$$\langle \varepsilon_n \rangle = \frac{1}{M_{\mathrm{ev}}} \sum_{j=1}^{M_{\mathrm{ev}}} \varepsilon_n \{\mathrm{event} \ j\}$$



- Quantifying closure:
 - Value of posterior at true value of parameters





- Quantifying closure:
 - Value of posterior at true value of parameters



- Akaike information criterion
 - $= -2 \ln \text{Likelihood}_{max} + 2 \text{ (number of model parameters)}$
 - Used for model comparison



- Quantifying closure:
 - Value of posterior at true value of parameters



- Akaike information criterion
 - $= -2 \ln L_{max} + 2$ (number of model parameters)
 - Used for model comparison

Best use of budget (best constraints) when $N_{param \ samples}/M_{event} \sim 0.1 - 1$

2 observables, 2 parameters (changing params)





2 observables, 2 params (changing uncert. of "data")





2 observables, 2 params (changing uncert. of "data")



"Data uncertainty" 2X smaller



2/3/4 observables, 2 params (changing # of observables)





$$\varepsilon_{n} e^{in\Phi_{n}} = \frac{\int_{0}^{\infty} drr \int_{0}^{2\pi} d\phi \ r^{n} \epsilon(r, \phi) e^{in\phi}}{\int_{0}^{\infty} drr \int_{0}^{2\pi} d\phi \ r^{n} \epsilon(r, \phi)}$$
$$\langle \varepsilon_{n} \rangle = \frac{1}{M_{ev}} \sum_{j=1}^{M_{ev}} \varepsilon_{n} \{\text{event } j\}$$

2/3/4 observables, 2 params (changing # of observables)

2 observables

3 observables

4 observables



2/3/4 observables, <u>3 params</u> (changing # of observables)





Best use of budget (best constraints) when $N_{param \, samples}/M_{event} \sim 0.1 - 1$ or $N_{param \, samples}^{optimal?} \approx 0.25 \cdot 1 \sqrt{M_{event} \times N_{param \, samples}}$ Budget



- Best use of budget (best constraints) when $N_{param \ samples}/M_{event} \sim 0.1 - 1$ or $N_{param \ samples}^{optimal?} \approx 0.25 \cdot 1 \sqrt{M_{event} \times N_{param \ samples}}$ Budget
- What had been used by contemporary publications?
 - $N_{param \ samples} \sim 10^3$
 - $M_{event} \sim 10^3 \cdot 10^5$ (some 10^6)
 - So $N_{param \ samples}/M_{event} \sim 0.01 1$

(overprioritizing statistical uncertainty over interpolation uncertainty?)

Model responses of an observable with respect to a given parameter

Analysis

Does this generalize?

Depends on:

- Accuracy of data
- Sensitivity of observables to parameters

collisions

2016



Summary

- Stochastic simulations have additional trade-offs when optimizing analyses
- Depends on constraints provided by data on different parameters given model
- We used a simple model to study trade-offs

$$N_{param \ samples}^{optimal?} / M_{event} \sim 0.1 - 1$$

 $N_{param\,samples}^{optimal?} \approx 0.25-1 \sqrt{budget}$



QUESTIONS?

PAPER • OPEN ACCESS

Computational budget optimization for Bayesian parameter estimation in heavy-ion collisions

Brandon Weiss¹ D, Jean-François Paquet^{1,2} D and Steffen A Bass¹ D Published 16 May 2023 • © 2023 The Author(s). Published by IOP Publishing Ltd Journal of Physics G: Nuclear and Particle Physics, Volume 50, Number 6 Citation Brandon Weiss *et al* 2023 *J. Phys. G: Nucl. Part. Phys.* 50 065104 DOI 10.1088/1361-6471/acd0c7

[arXiv:2301.08385]

BACKUP



Prior: example for the shear viscosity $\eta/s(T)$

- Positive definite; continuous function of temperature T (at zero chemical potential)
- Large values may be excluded by model self-consistency, causality, ...
- Theoretical constraints? Self-consistency across model stages?
- 100% C.I. (Prior) Guidance from other substances (minimum near crossover) 0.8 90% C.I. (Prior) 60% C.I. (Prior) 0.20 0.6 0.15 s/L S/L 0.10 $(\eta/s)_{kink}$ a_{low} 0.2 0.05 0.0 0.00 0.2 0.3 0.1 0.1 0.2 T_n 0.3 0.4 T (GeV) T[GeV]

0.4

Uncertainty optimization

Weiss et al (2023) [arXiv:2301.08385]

Model is stochastic (need to average over large number of collisions)



Emulation

posterior(
$$\overline{param}$$
) $\propto prior(\overline{param}) \times exp\left(-\frac{1}{2}\left(\overline{Model}(\overline{param}) - \overline{D}\right)^T \operatorname{Covar}^{-1}(\overline{Model}(\overline{param}) - \overline{D})\right)$

- Posterior is high-dimensional, and we cannot sample it easily for all values of the parameters
 - Option A: compute the posterior at a sample of model parameters and interpolate
 - Option B: compute the model's prediction at a sample of model parameters and interpolate



From impact geometry to momentum anisotropy



- Spatial anisotropy from partial overlap of nuclei & fluctuation
- Interactions transfer spatial anisotropy into momentum one
- Rapid development of momentum anisotropies consistent with strongly-coupled system



Nuclear Theory

[Submitted on 26 Oct 2023]

Applications of emulation and Bayesian methods in heavy-ion physics

Jean-François Paquet

Contents

1	Intr	roducti	ion	4	
	1.1	Paran	neter estimation in heavy-ion physics	4	
	1.2	Hard	and soft physics, and observable factorization in heavy-ion collisions	5	
	1.3	Struct	ure of this review	6	
2	Bri	ef ove	rview of emulation and Bayesian inference in heavy-ion		
	coll	isions		6	
	2.1	Model	and parameters	6	
	2.2	Measu	rements and observables	7	
	2.3	Model	-to-data comparison with Bayesian inference	8	
	2.4	Emula	tion	10	
3	Em	Emulation and applications to heavy-ion collisions			
	3.1	Model	emulation - overview	11	
	3.2	Dimer	nsionality reduction of the observables	13	
	3.3	Sampl	ing of the parameter space and emulation uncertainties	13	
		3.3.1	Prior and sampling of the parameter space	14	
		3.3.2	Balancing interpolation uncertainty and statistical uncertainty	15	
		3.3.3	Static and adaptive sampling	16	
	3.4	Emula	tion with Gaussian process regressors	18	
	3.5	Emula	tor validation	20	
	3.6	New d	levelopments in emulation	21	
		3.6.1	Transfer learning	21	
		3.6.2	Multifidelity with non-ordered models	22	
		3.6.3	Multifidelity with accuracy parameter extrapolation	23	
		3.6.4	Nonparametric approach for functional model parameters	23	
	3.7	Uses o	of model emulators	2 4	
		3.7.1	Sensitivity analysis and correlations	25	
		3.7.2	Visualization of parameter dependence of observables $\ldots \ldots \ldots$	25	

	Вау	/esian inference: general concepts	26		
	4.1	Bayesian inference	27		
		4.1.1 Priors	27		
		4.1.2 Likelihood	29		
	4.2	Numerical aspects	30		
	4.3	Maximum a posteriori parameters	31		
	4.4	Model selection	31		
	4.5	Model averaging and mixing	32		
	4.6	Validation with closure tests	33		
5	Bayesian inference: applications in heavy-ion physics				
	5.1	Overview	34		
			<u> </u>		
	5.2	Bayesian constraints on the viscosity	34		
6	5.2 Fut	Bayesian constraints on the viscosity	34 43		
6	5.2 Fut 6.1	Bayesian constraints on the viscosity	34 43 43		
6	5.2 Fut 6.1 6.2	Bayesian constraints on the viscosity	34 43 43 44		
6	5.2 Fut 6.1 6.2 6.3	Bayesian constraints on the viscosity	 34 43 43 44 45 		

Model-data comparison

Experimental uncertainties lead to uncertainties on the model parameters



Bayes' theorem

Experimental uncertainties lead to uncertainties on the model parameters

Constraints from Bayesian inference:

posterior(\overline{param}) $\propto prior(\overline{param}) \times exp\left(-\frac{1}{2}\left(\overline{Model}(\overline{param}) - \overline{D}\right)^T \operatorname{Covar}^{-1}\left(\overline{Model}(\overline{param}) - \overline{D}\right)\right)$

Bayes theorem:

$$prob(d) \times prob(p|d) = prob(p,d) = prob(p) \times prob(d|p)$$

 Evidence
 ×
 Posterior
 =
 Joint
 =
 Prior
 ×
 Likelihood

 [how likely are parameters given data]
 [how likely are data given parameters]
 [how likely are data given parameters]

Note: Bayes' theorem says nothing about choice of likelihood function

Bayes' theorem, prior and iterative constraints

Constraints from Bayesian inference:

posterior(\overline{param}) $\propto prior(\overline{param}) \times exp\left(-\frac{1}{2}\left(\overline{Model}(\overline{param}) - \overline{D}\right)^T \operatorname{Covar}^{-1}\left(\overline{Model}(\overline{param}) - \overline{D}\right)\right)$

- In theory: posterior from one Bayesian inference (with data set #1) becomes prior for the next (with data set #2)
- In practice:
 - Models are being improved
 - Re-use of previous posteriors has been rare

Note: prior should be independent of set of data currently being compared to

Model-data comparison

- Experimental uncertainties lead to uncertainties on the model parameters
- Constraints from Bayesian inference:

posterior(\overline{param}) $\propto prior(\overline{param}) \times exp\left(-\frac{1}{2}\left(\overline{Model}(\overline{param}) - \overline{D}\right)^T \operatorname{Covar}^{-1}\left(\overline{Model}(\overline{param}) - \overline{D}\right)\right)$

- Posterior has the dimension of the number of parameters
- Marginalized posterior: integrating posterior over all parameters except "n"



Model-data comparison

Constraints from Bayesian inference:

posterior(\overline{param}) $\propto prior(\overline{param}) \times exp\left(-\frac{1}{2}\left(\overline{Model}(\overline{param}) - \overline{D}\right)^T \operatorname{Covar}^{-1}\left(\overline{Model}(\overline{param}) - \overline{D}\right)^T\right)$

Marginalized posterior: integrating posterior over all parameters except 1 or 2 or ...

Marg.posterior $\left(\frac{\eta}{s}\right) = \int d(initial \ cond. parameters) d(bulk \ viscosity \ param) d(...)$ posterior $\left(\frac{param}{param}\right)$



Different analyses = different constraints

- Use different data sets
- Different modelling assumptions:
 - Hydrodynamics
 - Initial conditions
 - Cooper-Frye
 - Parameters and priors
- Treatment of correlations in experimental uncertainties

Experimental uncertainties and covariance matrix posterior(\overline{param}) $\propto prior(\overline{param}) \times exp\left(-\frac{1}{2}\left(\overline{Model}(\overline{param}) - \overrightarrow{D}\right)^T \operatorname{Covar}^{-1}\left(\overline{Model}(\overline{param}) - \overrightarrow{D}\right)\right)$ $\left[\left(y_1(\vec{p}) - y_1^{expt} \right) \quad (y_2(\vec{p}) - y_2^{expt}) \right] (\text{Covariance matrix})^{-1} \begin{bmatrix} (y_1(\vec{p}) - y_1^{expt}) \\ (y_2(\vec{p}) - y_2^{expt}) \end{bmatrix}$ $Cov = \begin{bmatrix} (\sigma_1^{expt})^2 & 0 \\ 0 & (\sigma_2^{expt})^2 \end{bmatrix} \xrightarrow{\begin{subarray}{c} 1.0 \\ 0.8 \\ 0.8 \\ 0.6 \\ 0.6 \\ 0.4 \\ 0. \end{subarray}$ Uncorrelated uncertainties: (stat. uncert.?) Fully-correlated uncertainties: (normalization uncert.?) 0.2 Partly-correlated uncertainties: $\text{Cov} = \begin{bmatrix} (\sigma_1^{expt})^2 & Cov(1,2) \\ Cov(2,1) & (\sigma_2^{expt})^2 \end{bmatrix}$ 50 (systematic uncert.?) Centrality (%)

Uncertainties and covariance matrix posterior(\overline{param}) $\propto prior(\overline{param}) \times exp\left(-\frac{1}{2}\left(\overline{Model}(\overline{param}) - \overline{D}\right)^T \operatorname{Covar}^{-1}\left(\overline{Model}(\overline{param}) - \overline{D}\right)\right)$ $\left[\left(y_1(\vec{p}) - y_1^{expt} \right) \quad (y_2(\vec{p}) - y_2^{expt}) \right] (\text{Covariance matrix})^{-1} \begin{bmatrix} (y_1(\vec{p}) - y_1^{expt}) \\ (y_2(\vec{p}) - y_2^{expt}) \end{bmatrix}$ $\begin{aligned} \text{Covariance matrix} &= \begin{bmatrix} \left(\sigma_{1}^{expt,uncorr}\right)^{2} & 0 \\ 0 & \left(\sigma_{2}^{expt,uncorr}\right)^{2} \end{bmatrix} + \left(\sigma^{expt,fully\,corr}\right)^{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} + \\ \begin{bmatrix} \left(\sigma_{1}^{expt,corr}\right)^{2} & cov(1,2) \\ cov(2,1) & \left(\sigma_{2}^{expt,corr}\right)^{2} \end{bmatrix} + \end{aligned}$ (emulator covariance)+(model statistical uncertainty)

Hydrodynamic-based simulations of heavy ion collisions

Nch / (Nch)

Successful in describing broad sets of measurements

Nijs, van der Schee, Gürsoy, Snellings (2021) PRC, PRL - π^{\pm} PbPb, $\sqrt{s_{NN}}$ =2.76 TeV 0.20 PbPb, √ s_{NN} =2.76 TeV 1000 Kt $\frac{dN/N_{ev}}{dp_{T} dy} [GeV^{-1}c]$ 0.15 100 {z} ^2 0.10 10 0.05 0.1 0.00 0.5 1.0 1.5 2.0 2.5 3.0 0.5 1.0 1.5 2.0 pT [GeV/c] pT [GeV/c] 0.10 π^{\pm} pPb, $\sqrt{s_{NN}} = 5.02 \text{ Te}$ 1.4 pPb, √s_{NN} =5.02 TeV 0.08 0.06 (p₁) [GeV/c] $\tilde{V}_{n(k)}$ 0.04 0.02 0.00 0.6 $\tilde{v}_{2}\{2\} = \tilde{v}_{3}\{2\}$ -0.02 0.4 10 20 30 40 50 60 2 0 3

centrality [%]



Interaction and expansion



