Brigham Young University
BYU ScholarsArchive

Theses and Dissertations

2021-12-20

Bayesian, Frequentist, and Information Geometry Approaches to Parametric Uncertainty Quantification of Classical Empirical Interatomic Potentials

Yonatan Kurniawan
Brigham Young University

Follow this and additional works at: https://scholarsarchive.byu.edu/etd

 Part of the Physical Sciences and Mathematics Commons

BYU ScholarsArchive Citation

Kurniawan, Yonatan, "Bayesian, Frequentist, and Information Geometry Approaches to Parametric Uncertainty Quantification of Classical Empirical Interatomic Potentials" (2021). *Theses and Dissertations*. 9819.
https://scholarsarchive.byu.edu/etd/9819

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact ellen_amatangelo@byu.edu.

Bayesian, Frequentist, and Information Geometry Approaches to Parametric Uncertainty

Quantification of Classical Empirical Interatomic Potentials

Yonatan Kurniawan

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Mark Transtrum, Chair
Gus Hart
Dennis Della Corte

Department of Physics and Astronomy

Brigham Young University

Copyright © 2021 Yonatan Kurniawan

All Rights Reserved

ABSTRACT

Bayesian, Frequentist, and Information Geometry Approaches to Parametric Uncertainty Quantification of Classical Empirical Interatomic Potentials

Yonatan Kurniawan
Department of Physics and Astronomy, BYU
Master of Science

Uncertainty quantification (UQ) is an increasingly important part of materials modeling. In this paper, we consider the problem of quantifying parametric uncertainty in classical empirical interatomic potentials (IPs). Previous work based on local sensitivity analysis using the Fisher Information has shown that IPs are sloppy, i.e., are insensitive to coordinated changes of many parameter combinations. We confirm these results and further explore the non-local statistics in the context of sloppy model analysis using both Bayesian (MCMC) and Frequentist (profile likelihood) methods. We interface these tools with the Knowledgebase of Interatomic Models (OpenKIM) and study three models based on the Lennard-Jones, Morse, and Stillinger-Weber potentials, respectively. We confirm that IPs have global properties similar to those of sloppy models from fields such as systems biology, power systems, and critical phenomena. These models exhibit a low effective dimensionality in which many of the parameters are unidentifiable, i.e., do not encode any information when fit to data. Because the inverse problem in such models is ill-conditioned, unidentifiable parameters present challenges for traditional statistical methods. In the Bayesian approach, Monte Carlo samples can depend on the choice of prior in subtle ways. In particular, they often “evaporate" parameters into high-entropy, sub-optimal regions of the parameter space. For profile likelihoods, confidence regions are extremely sensitive to the choice of confidence level. To get a better picture of the relationship between data and parametric uncertainty, we sample the Bayesian posterior at several sampling temperatures and compare the results with those of Frequentist analyses. In analogy to statistical mechanics, we classify samples as either energy-dominated, i.e., characterized by identifiable parameters in constrained (ground state) regions of parameter space, or entropy-dominated, i.e., characterized by unidentifiable (evaporated) parameters. We complement these two pictures with information geometry to illuminate the underlying cause of this phenomenon. In this approach, a parameterized model is interpreted as a manifold embedded in the space of possible data with parameters as coordinates. We calculate geodesics on the model manifold and find that IPs, like other sloppy models, have bounded manifolds with a hierarchy of widths, leading to low effective dimensionality in the model. We show how information geometry can motivate new, natural parameterizations that improve the stability and interpretation of UQ analysis and further suggest simplified, less-sloppy models.

Keywords: interatomic models, uncertainty quantification, sloppy models, Bayesian inference, profile likelihood, information geometry

ACKNOWLEDGMENTS

I would like to especially thank Mark Transtrum for his patience and guidance as we work on this challenging problem. He continually encouraged me to think critically and helped me to build understanding of this problem. He always motivated me to move forward and assisted me to grow academically.

I would also like to thank BYU KIM group, especially Cody Petrie and Kinamo Williams. I want to acknowledge their contributions to this project as well as in the development of the interface we used. Additionally, they also helped me in many different ways, such as providing feedback on the program I have written and having independent group discussions.

To my family who always supports me during my journey. I would especially thank my wife who always supports and loves me in every time. To my brother, who always sets an example in our family in pursuing a higher education.

Finally, I would like to especially thank OpenKIM and its developers. OpenKIM had provided the materials we used in this project. The developers, especially Ellad Tadmor, Ryan Elliott, Daniel Karls, and Mingjian Wen had helped me with the computational setup and provided me with their guidance.

This work is supported by the U.S. National Science Foundation under award CMMT-1834332. Resource at BYU Office of Research Computing is greatly appreciated.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Interatomic potentials (IPs) are a foundational tool in computational materials science [1]. They allow modelers to make efficient predictions of materials properties without reference to the complicated sub-atomic structure. Recently there has been considerable interest in applying methods of uncertainty quantification (UQ) to IPs [2–5]. UQ assesses the reliability of materials predictions, leveraging tools from statistical inference [6]. Statistical analysis of similar inverse problems in physics has motivated the study of *sloppy models* [7]. Sloppy models lead to extremely ill-conditioned inverse problems and pose several challenges for standard statistical methods [8, 9]. This work considers the application of UQ to IPs in the context of sloppy models. We find that many IPs are sloppy, which leads to challenges in interpreting UQ results, and use information geometry to mitigate some of these challenges.

Classical IPs have been widely used in materials science to circumvent the computational cost of quantum calculations, such as density functional theory (DFT), by approximating the interaction energy between atoms without considering the electrons. However, a definitive functional form that can describes all types of atomic bonding hasn't been discovered. Thus, IPs are often designed for specific purposes, resulting in a plethora of models [10]. Efforts such as the Open Knowledgebase of Interatomic Models (OpenKIM) [11] aim to organize and standardize these IPs.

In the development of classical empirical IPs, the parameters are typically fit to match experimental or first principles data of some microscopic properties such as the lattice parameters and elastic constants of single crystals, or the potential energy and atomic forces associated with random atomic configurations [12]. They are then used in conjunction with simulation codes to predict other properties that are not used in the fitting process. Thus, UQ is relevant for assessing the reliability of these out-of-sample predictions.

In this paper, we are primarily interested in parametric uncertainty, i.e., uncertainty in the model's parameters, which is quantified through, for example, a Bayesian posterior distribution or confidence regions on the parameter space. In our formulation, the IP is used in two models, one that makes predictions for training data (e.g., energy and forces) and a second that makes predictions for other material quantities of interest. By varying the parameters, the IP traces out a set of possible predictions for each model. The set of predictions made by a model is known as the *model manifold* [8] and studied by information geometry. Parameters act as coordinates on the model manifold and distances on the manifold measure statistical distinguishability. The UQ process propagates uncertainties from training data to uncertainties in parameters via the (pseudo) inverse of the first model. These uncertainties then propagate through the second model to give uncertainties in the predictions for the quantities of interest. The inverse of the model is not given explicitly but only accessible through iterative evaluations of the model. Consequently, the first uncertainty propagation process is the more challenging one and the focus of this study. Given the parametric uncertainties, they can then be propagated to other quantities of interest, for example as in [13]. This entire process is illustrated in Fig. 1.1.

Many UQ methods have been developed to propagate uncertainties in data to uncertainties in parameters. In materials science, Markov Chain Monte Carlo (MCMC) sampling of the Bayesian posterior is the most common approach. Being a Bayesian method, this requires a prior distribution, and several prior distributions have been used, including uniform [14–20], normal [21], Jeffreys

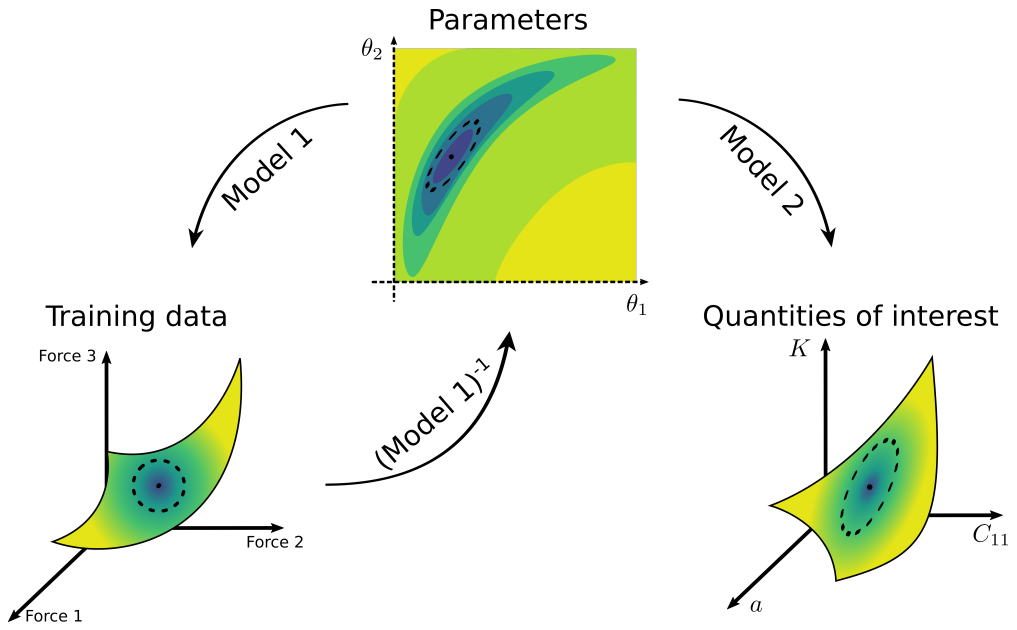


Figure 1.1 Uncertainty propagation in information geometry. An IP is described by several parameters, collectively forming a parameter space (top center). By considering the predictions for all allowed parameter values, a model maps out a set of possible predictions in data space, known as the model manifold (bottom left, right). Uncertainty in training data (bottom left) are propagated through the inverse function to uncertainties in parameter space, represented here by contours of constant likelihood, i.e., cost, in the parameter space. Parametric uncertainty can then be propagated forward to material quantities of interest, such as lattice constant, elastic constant, and bulk modulus (bottom right).

prior [22], and maximum entropy [23]. Other approaches to UQ include: F-statistics estimations [24], ANOVA-based methods [25], and multi-objective optimization [26]. Other fields have used the profile likelihood method [27–29], but to the best of our knowledge has not yet been applied to IPs.

The study of inverse problems in statistical physics has identified an important property of many multi-parameter models, known as *sloppiness*. Sloppy models are characterized by predictions that are insensitive to coordinated changes in combinations of parameters. Inverse problems for sloppy models are extremely ill-conditioned and, as we will show below, present obstacles for standard UQ methods. Sloppiness was first systematically studied in 2003 by Brown and Sethna in the context of systems biology models [7]. The relevant object is the Fisher Information Matrix (FIM) [30] that quantifies the information that data carry about parameters in a model. Eigenvalues of the FIM provide a local measure of sloppiness. For sloppy models, the FIM eigenvalues span many orders of magnitude and have many small eigenvalues. These small eigenvalues correspond to sloppy combinations of parameters, i.e., those that are ill-constrained by data [8, 31–33].

Although sloppy models are usually identified by their characteristic FIM spectrum, the theory of sloppy models is couched in the tools of information geometry, the application of differential geometry to statistics [9, 34]. The key object is the model manifold, which comprises the set of all possible predictions a model can make with different parameter values. As we show in Sec. 2.5 the model manifold is embedded in data space. This space is useful because distance corresponds to statistical identifiability. In other words, points that are distant on the model manifold are statistically distinguishable, while nearby points are not. Statistical identifiability therefore induces a Riemannian metric on the parameter space that is given by the FIM. For sloppy models, the model manifold is bounded with a hierarchy of widths, suggesting the model exhibits a low effective dimensionality [8]. When a model manifold is very thin, the parameters associated with the thin directions are unidentifiable from data.

Sloppy models are ubiquitous in many scientific fields including critical phenomenon [35], systems biology [9, 31, 36–39], power systems stability [40], particle accelerators [41], and others [33]. In molecular modeling, it has been shown that IPs typically exhibit the same characteristic sloppy FIM spectrum [2–4, 42]; however, the techniques of information geometry have not yet found application to IPs. As materials models increase in complexity, especially with the advent of machine learning models [5, 43–45], sloppiness will become increasingly relevant. This underscores the importance of understanding the effects of sloppiness on IPs and other materials science models.

This work considers the role of sloppiness for the application of UQ to IPs. Since sloppy models are ill-conditioned, they pose unique challenges for standard UQ methods. Results are often highly sensitive to details of the problem formulation and difficult to interpret. We illustrate these challenges with both Bayesian and Frequentist UQ methods, introducing the profile likelihood methods to the material modeling community. We additionally bring the information geometry techniques of sloppy model analysis to bear on this problem and discuss how they can illuminate and mitigate these challenges. The paper is organized as follows. First, we precisely formulate the problem in Sec. 2.1 and discuss using the FIM as a local analysis of sloppiness in Sec. 2.2. Then, we discuss the Bayesian and Frequentist UQ methods in Sec. 2.3 and 2.4, respectively. We describe information geometric tools, specifically geodesics, central to our sloppy model analysis in Sec. 2.5. Sec. 2.6 presents the models used in this study, i.e., the IPs and quantities of interest. We present the results for each method in Ch. 3. Finally, we discuss the effects sloppiness has on UQ in Ch. 4 and the prospect for accurate, efficient UQ in IPs generally.

Chapter 2

Methods

In this chapter we introduce the general methods that will be used later on in the thesis. We pay particular attention to the mathematical assumptions that different methods require and the types of calculations that they enable. Then we introduce the interatomic models and data sets on which we conduct our study.

2.1 Defining Cost

The minimal elements for parametric UQ are (1) a collection of data, $\{y_m\}_{m=1}^M$ (where M is the number of data points), (2) a parameterized family of models that make predictions $\{f_m(\theta)\}_{m=1}^M$, and (3) a metric for comparing the model predictions to data, $\|\cdot\|$. We assume that the model depends on N parameters $\theta \in \mathcal{D} \subseteq \mathbb{R}^N$. Here, $\mathcal{D}$ is the physically allowed domain; for example, it is common for some parameters to be restricted to positive values. It is convenient to interpret both the data and model predictions as vectors in an M-dimensional *data space*: $y_m \to \mathbf{y} \in \mathbb{R}^M$, $f_m(\theta) \to \mathbf{f} : \mathcal{D} \subseteq \mathbb{R}^N \to \mathbb{R}^M$.

The third requirement, a metric, defines a *cost function* (also known as a loss function) that quantifies how well specific parameter values fit the available data, $C(\theta) = \|\mathbf{y} - \mathbf{f}(\theta)\|$. The best

fit parameters, denoted by θ^*, minimize the cost. By far the most common choice for a metric is (weighted) least squares

$$C(\theta) = \frac{1}{2}\sum_{m=1}^{M} r_m(\theta)^2, \tag{2.1}$$

where we have introduced the *residuals*

$$r_m(\theta) = \frac{y_m - f_m(\theta)}{\sigma_m} \tag{2.2}$$

that depends on the inverse weights σ_m that act as error bars for each data point.

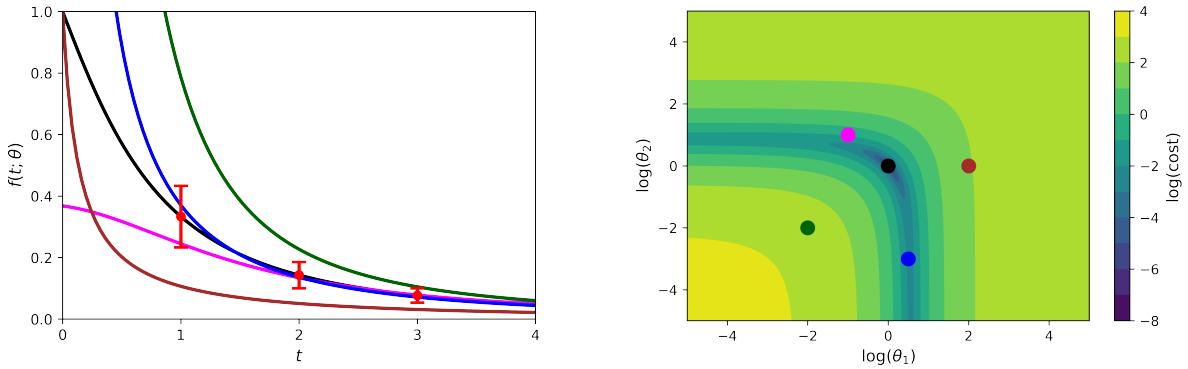
The cost function has a probabilistic interpretation as the negative log-likelihood

$$P(y|\theta) \sim \exp\{-C(\theta)\}. \tag{2.3}$$

Eq. 2.1 corresponds to the case that residuals are independent and identically distributed Gaussian random variables: $r_m \sim \mathcal{N}(0,1)$, or equivalently $y_m \sim \mathcal{N}(f_m(\theta), \sigma_m^2)$. Probability acts as a *measure* on data space that we use to quantify uncertainty. For stochastic processes, the stochastic variation in the data is a natural measure, in which case the inverse weights, σ_m, are often taken to be the standard errors estimated from repeated observations. When working with DFT data, we often do not have an associated error bar. Instead, we advocate selecting σ_m to be the acceptable tolerance in the model for a particular application. Deliberate selection of σ_m is an important part of the UQ problem formulation since any eventual measure of the uncertainty in the model parameters will be derived from the choice of measure in data space. Because predictions may be made for quantities that carry different physical units (e.g., energies vs. forces), choosing σ_m is minimally necessary for Eq. 2.1 to be dimensionally consistent. In general, choosing σ_m as a fractional tolerance, e.g., 10% of the data, is a reasonable choice and what we use in this study.

To illustrate key ideas throughout this section, we use a two-parameter toy model of the form

$$f(t;\theta) = \frac{1}{t^2 + \theta_1 t + \theta_2}. \tag{2.4}$$



(a) **(b)**

Figure 2.1 Time series (a) and cost surface (b) for the toy model in Eq. 2.4. Data are given by the red points in (a), with the error bars set to 30% of the data. Model predictions for different parameter values and their corresponding points on the cost surface are shown in matching colors. Contours represent the errors between the model predictions and data as calculated in Eq. 2.1.

We make predictions at times $t = 1.0, 2.0, 3.0$, i.e., $f_m(\theta) = f(t_m; \theta)$. We restrict $\theta_i \geq 0$ (a common physical constraint on parameter values), which suggests working with the log-transformed parameter values. We use $\mathbf{y} = [1/3, 1/7, 1/13]^T$ with tolerances σ_m set to be 30% of the data $\mathbf{y}$, for the purpose of visual clarity. This data, along with model predictions for several values of the parameters, are shown in Fig. 2.1a.

It is useful to visually consider cost contours for this model, shown in Fig. 2.1b. In general, we are interested in describing the regions in parameter space with low cost. The model in Eq. 2.4 is sloppy, as manifested by its insensitivity to coordinated variations in some parameter directions. Because of this, the cost contours in the sloppy directions are elongated and the aspect ratio of the canyon around the minimum is very large. Away from the best fit, many of the cost contours do not close; canyons stretch to the edges of parameter space and flatten into broad plateaus. Extreme values of the parameters can have finite, and in many cases, very small cost. These features are ubiquitous in sloppy models, and play a fundamental role in quantifying the parametric uncertainty.

For models with many parameters, direct visualization of the cost surface is not possible. However, we have constructed this toy model to illustrate issues that are typical of high-dimensional parameter spaces. In the next sections, we describe several tools for analyzing the cost surface of multi-parameter models.

2.2 Fisher Information: Sloppy Model Analysis

To quantify the local geometry of the cost surface in a neighborhood of the best fit, we linearize the residuals about θ^*:

$$r_m(\theta) \approx r_m(\theta^*) + \frac{\partial r_m}{\partial \theta}(\theta - \theta^*). \tag{2.5}$$

To lowest order, the cost function becomes

$$C(\theta) \approx C(\theta^*) + \frac{1}{2}(\theta - \theta^*)^T (J^T J)(\theta - \theta^*), \tag{2.6}$$

where we have used the fact that $\nabla C = 0$ at θ^* and introduced the Jacobian of the residual function $J_{mn} = \partial r_m / \partial \theta_n = -(1/\sigma_m)\partial f_m / \partial \theta_n$ evaluated at θ^* . The squared Jacobian appearing in Eq. 2.6 is the Fisher Information Matrix (FIM):

$$\mathcal{I} = J^T J. \tag{2.7}$$

The FIM is an important statistical quantity; its inverse is a lower bound on the covariance of parameter uncertainty, known as the the Cramér-Rao bound [46].

The local geometry of the cost surface around the best fit is described by the FIM, as we illustrate in Fig. 2.2. Diagonals of the FIM describe the change in cost to each parameter individually, ignoring any potential correlations among parameters. Cost contours form ellipses, aligned with the eigenvectors of the FIM, whose aspect ratio is given by the square root of the ratio of the eigenvalues. Elongated directions are parallel to the eigenvectors with small eigenvalues, indicating that the data carry little information about those parameter combinations. These parameter combinations are

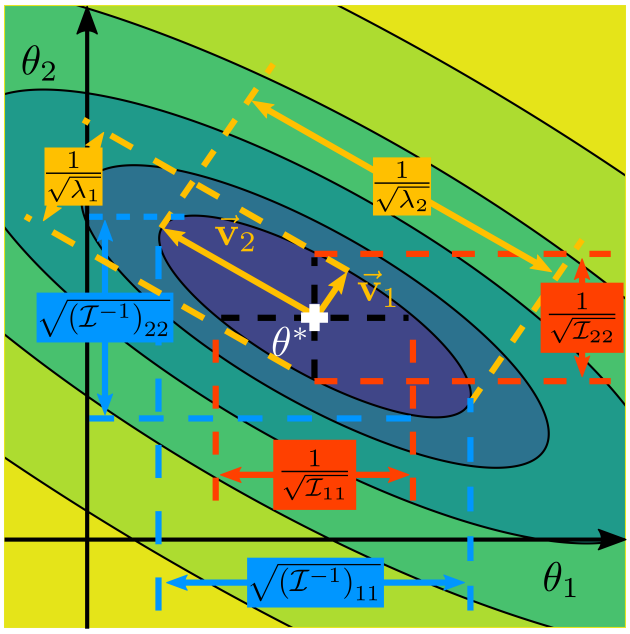


Figure 2.2 The Fisher Information describes the local geometry of the cost surface. Locally cost contours are represented by ellipses where the axes are aligned with the eigenvectors of the FIM. The aspect ratio of the ellipse is given by the square root of the ratio of the eigenvalues.

only weakly constrained by the data and have large uncertainties in their inferred values. Projecting these ellipses onto the parameter axes estimates the uncertainty in each individual parameter, given by the diagonals of the inverse FIM.

2.3 Bayesian Analysis

The most common UQ methods in molecular modeling use a Bayesian framework. In Bayesian statistics, the parametric uncertainty is described by a posterior distribution given by Bayes' theorem

$$P(\theta|\mathbf{y}) \propto L(\theta|\mathbf{y}) \cdot \pi(\theta), \tag{2.8}$$

where $L(\theta|\mathbf{y})$ and $\pi(\theta)$ are the likelihood and the prior distribution of the model's parameters, respectively [47, 48]. The likelihood is functionally the same as the probability distribution of

the observed data conditioned on the different values of the parameters, i.e., $L(\theta|\mathbf{y}) = P(\mathbf{y}|\theta)$. A formal analogy to the Boltzmann distribution [7, 49, 50] suggests the introduction of a "sampling temperature," T, which modifies Eq. 2.3 as

$$L(\theta|\mathbf{y}) \sim \exp\{-C(\theta)/T\}. \tag{2.9}$$

The cost is analogous to the internal energy of a system, so low temperature distributions are concentrated near the low-energy (i.e., low cost) region of parameter space. Formally, the temperature uniformly scales the tolerances σ_m in Eq. 2.2. Continuing the analogy, Eq. (2.8) becomes

$$P(\theta|\mathbf{y}) \sim \exp\left(-\left(C(\theta) - S(\theta)T\right)/T\right), \tag{2.10}$$

so the prior is analogous to entropy: $S = \log \pi$.

We sample from the posterior distribution using an MCMC algorithm. There are several black-box libraries for MCMC sampling. In this work, we used the `ptemcee` Python package, which utilizes an affine invariance property of the sampler [51, 52]. In addition, this method generates chains at different temperatures and mixes them with an appropriate acceptance probability [53]. Parallel tempering improves convergence rates by allowing walkers to skip over regions in parameter space with higher cost values and possibly find different minima (if they exist). Additionally, by sampling at multiple temperatures, we can assess how the choice of σ_m affects any conclusions we draw from the distribution.

To assess the convergence, we simulate multiple chains and use the Potential Scale Reduction Factor (PSRF), denoted by $\hat{R}$ [54–56]. The value of $\hat{R}$ is related to the ratio of the covariance between and within the independent chains, given by

$$\hat{R} = \frac{n-1}{n} + \frac{m+1}{m}\lambda_{\max}(W^{-1}B/n), \tag{2.11}$$

where n and m are the numbers of iterations and chains, respectively, and $\lambda_{\max}(A)$ denotes the largest eigenvalue of matrix A. B/n and W are the variance between and within the independent

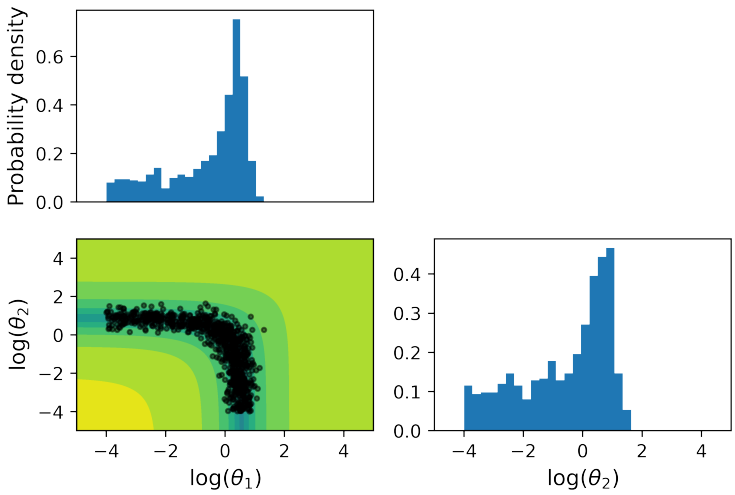


Figure 2.3 MCMC samples of the model defined by Eq. 2.4. On the lower triangle, the samples are plotted on top of the cost contour. The univariate marginal distributions of the samples are given on the diagonal.

chains, ψ_j,

$$\begin{aligned} \frac{B}{n} &= \frac{1}{m-1}\sum_{j=1}^{m}\left(\bar{\psi}_j - \bar{\psi}\right)\left(\bar{\psi}_j - \bar{\psi}\right)^T \\ W &= \frac{1}{m(n-1)}\sum_{j=1}^{m}\sum_{t=1}^{n}\left(\psi_{jt} - \bar{\psi}_j\right)\left(\psi_{jt} - \bar{\psi}_j\right)^T. \end{aligned} \tag{2.12}$$

As the MCMC samples converge to a stationary distribution, the value of $\hat{R}$ approaches one; however, the converse is not necessarily true. Common thresholds of $\hat{R}$ are in the range of 1.1 to 1.05 [55]. In this work, we have used the more stringent requirement ($\hat{R} < 1.05$).

To illustrate, we sample the posterior of the model in Eq. 2.4 with a uniform prior distribution that is non-zero over $(-4,4)$. Fig. 2.3 shows the result of the sampling as an array of plots that summarize the sample. Along the main diagonal, we plot the univariate marginal distributions, i.e., projection of the samples' distribution onto a parameter axis. In the lower triangle of the array we show samples in two dimensional parameter space. In higher dimensions, these plots are two-dimensional marginal distributions. In this example, we have superimposed the samples on top of the cost contours (compare this to Fig. 2.1b).

In Fig. 2.3, notice that the MCMC samples are located along the canyon. This is expected as the posterior is large in regions of higher likelihood, i.e., lower cost. Thus, the samples accurately conform to the cost contours and quantify the uncertainty in the parameter estimates. However, by inspecting Fig. 2.3, we can anticipate a potential problem when the cost contours have flat, elongated canyons that extend to extreme parameter values. In this scenario, it will be common to "evaporate" parameters, i.e., have samples that extend over the full range of model parameters. Parameter evaporation was first observed when sampling posterior distributions for sloppy models in systems biology [41]; however, the phenomenon occurs in molecular models as well, as we document below.

Closely related to the parameter evaporation is the question of choosing the prior distribution, $\pi(\theta)$, on parameter space. Here, we have used a uniform prior, a common choice for an uninformative prior. However, we find the prior can strongly influence the posterior distribution in obscure ways. Even the apparently innocuous uniform prior can introduce strong biases. Notice that in Fig. 2.3, the samples stop evaporating due to the boundaries of the prior distribution. Consequently, the marginal distribution of the samples have a hard cutoff at this boundary. With a broader prior, the posterior will be even wider, and samples may no longer be concentrated near the best fit. The long canyons and broad plateaus of the cost surface would dominate the samples. We intuitively explain this effect in terms of a trade-off between energy and entropy in the sampling process. While the single most probable parameter value is the best fit (i.e., ground state), there are many more sub-optimal parameter values along the canyon. In other words, the broad prior introduces a large entropy in some regions of the parameter space. For broad priors, the entropic contribution dominates the sampling.

The effect is clearly demonstrated with a simple example. Consider a one-dimensional cost function given by

$$C(\theta) = \begin{cases} C_0 & \theta < l \\ C_0 + \Delta & \theta > l \end{cases} \tag{2.13}$$

where θ is a non-negative parameter and C_0, Δ and l are non-negative constants. We take a prior $\pi(\theta) = \mathcal{U}(0, L)$, i.e., a uniform, flat prior from zero to a positive value of L. After calculating the posterior distribution, we find the average cost for this scenario to be

$$\langle C \rangle = \begin{cases} C_0 & L < l \\ C_0 + \Delta \left(\frac{L-l}{L+l\left(e^{\Delta/T}-1\right)} \right) & L > l. \end{cases} \tag{2.14}$$

Notice that for very large L (i.e., very broad prior), $\langle C \rangle \to C_0 + \Delta$. That is to say, for a sufficiently broad prior, the posterior distribution is dominated by bad fits (large cost) because of their large entropic contribution. This result holds for any non-zero sampling temperature and regardless of how bad the fit is (i.e., the size of Δ).

As we will see in Ch. 3, the trade-off between entropy and energy is even more nuanced for "sloppy" cost landscapes with many dimensions. Rarely is there an objectively "correct" prior, and any choice is almost certain to introduce artifacts into the statistics of the posterior. In these cases, it is unclear to what extent the posterior accurately reflects the target uncertainty. One solution is to sample with multiple priors and temperatures, a computationally expensive task, and try to assess the effect of prior and sampling temperature on the results. Because this practice generates multiple posteriors, it potentially undermines the Bayesian paradigm in which a single posterior summarizes all the information one has about the parameters of a model. However, we believe these extra steps are an important intermediate analysis in understanding the effect of the prior on the posterior and necessary for constructing a reliable posterior. Alternatively, one could use a formalism that does not require an a priori measure on parameter space. This is the domain of Frequentist statistics, which we discuss next.

2.4 Frequentist Analysis

Although much has been said about the philosophical differences between Bayesian and Frequentists [57, 58], here we use a functional distinction. The choice of prior in the previous section was a central question. The prior acts as a measure on parameter space, i.e., a weight function whose integral generalizes the concept of lengths and volumes [59]. In the frequentist approach no such measure exists. Without a measure on parameter space, we lose the machinery of a posterior distribution, but we also need fewer mathematical assumptions. With no prior, the goal is to describe the set of parameter values that have small cost (i.e., below some statistically defined threshold) without attaching any (probabilistic) weight to regions of parameter space. The tool we use to accomplish this is the profile likelihood [60].

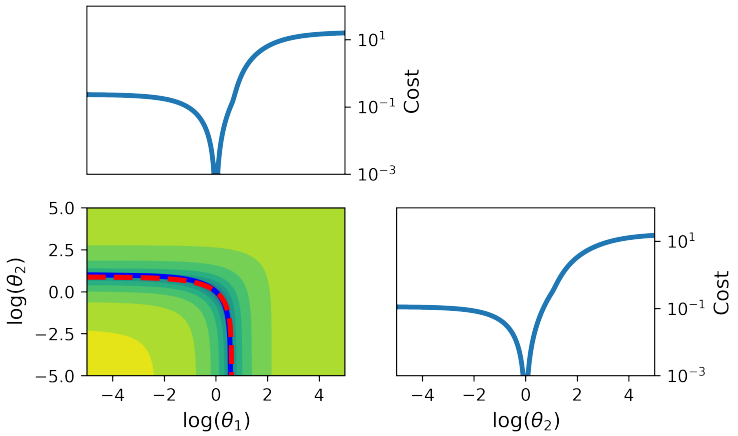


Figure 2.4 Profile likelihood for the model in Eq. 2.4. On the lower triangle, the red and blue curves show the paths traced from the profile likelihood computation for the parameter on the horizontal and vertical axes, respectively. The cost profiles, i.e., the cost along these paths, are given on the diagonal.

The basic idea is to select one parameter, fix it to a constant value, and globally optimize the likelihood function (i.e., minimize the cost) over the remaining $N-1$ model parameters [60, 61]. By varying the value to which the parameter is fixed, we trace out a "profile" of how the cost depends on this parameter in the context of the rest of the model. The procedure is best understood through example, as we now demonstrate on the toy model from Eq. 2.4 in Fig. 2.4. As before, we summarize results with a two-by-two array of figures. Consider the cost contours in the lower triangle of the plot array. The red curve is the set of points obtained by fixing θ_1 to a constant value and optimizing the cost over θ_2. The optimization searched over vertical slices of the parameter space for each value of θ_1. Similarly, the blue curve is the set of points found by fixing θ_2 and optimizing over horizontal slices, i.e., over θ_1. Along the main diagonal, we plot the cost along each of these profile likelihood paths.

By construction, the profile likelihood paths trace out the canyon on the cost contour. By comparing the paths and the cost along the paths, we extract information about how variation of the parameters affects the variation in the predictions. The paths also tell us how the parameters

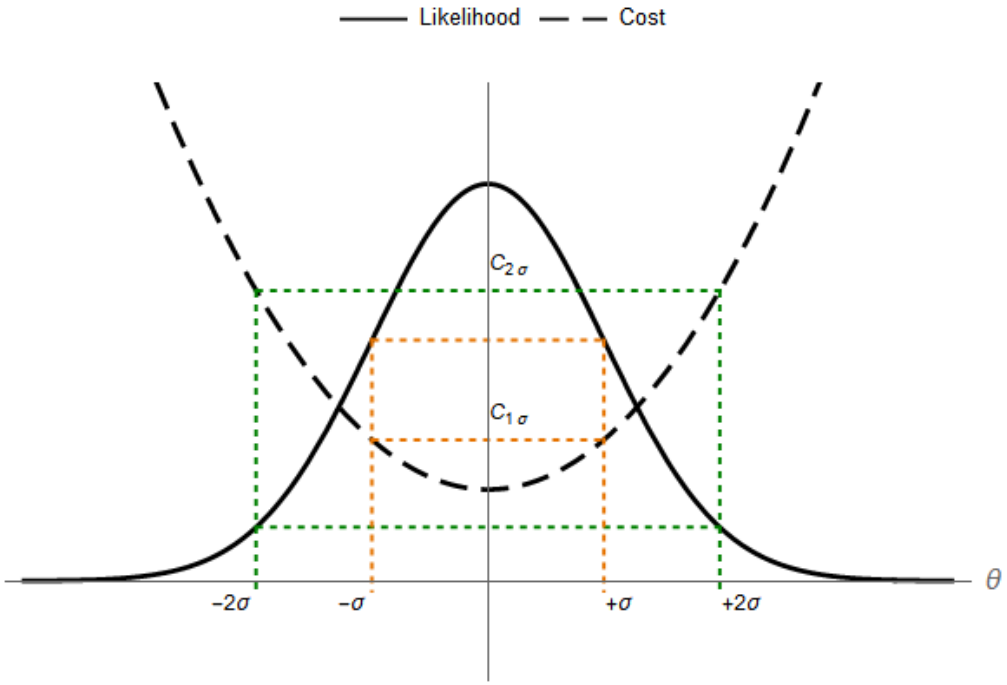


Figure 2.5 Confidence intervals for a Gaussian likelihood at several confidence levels. $C_{1\sigma}$ and $C_{2\sigma}$ correspond to the cost threshold at 68% and 95% confidence levels, respectively. The confidence interval of θ given 68% confidence level spans from $-\sigma$ to $+\sigma$.

correlate with each other. Statistical confidence levels correspond to an allowed error or cost threshold. For a given confidence level, the uncertainty of an individual parameter is given by the width of the profile likelihood that has cost values lower than this cost threshold. This idea is illustrated in Fig. 2.5 for a simple Gaussian likelihood function.

To calculate profile likelihoods, we developed a Python package, `profile_likelihood` [62] that additionally interfaces with IPs from the OpenKIM database at https://openkim.org. We use the Levenberg-Marquardt algorithm with geodesic acceleration in the optimization process [63].

The profile likelihood analysis method has its own challenges and limitations. Optimizing multi-dimensional cost functions can be challenging [8]; however, by using the result of the previous optimization as the starting point of each iteration, convergence is relatively fast and stable. Additionally, the profiling process effectively projects parameter curves onto the parameter axes. As

we will see in Ch. 3, if the cost canyon curves or bends, it will be missed by the profile likelihood. More broadly, the construction is not invariant to reparameterization. Just as priors may introduce artifacts in the Bayesian framework, the parameterization can introduce artifacts into the profile likelihood. To avoid these issues, we next use information geometry to study the uncertainty in a parameterization-independent way.

2.5 Information Geometry

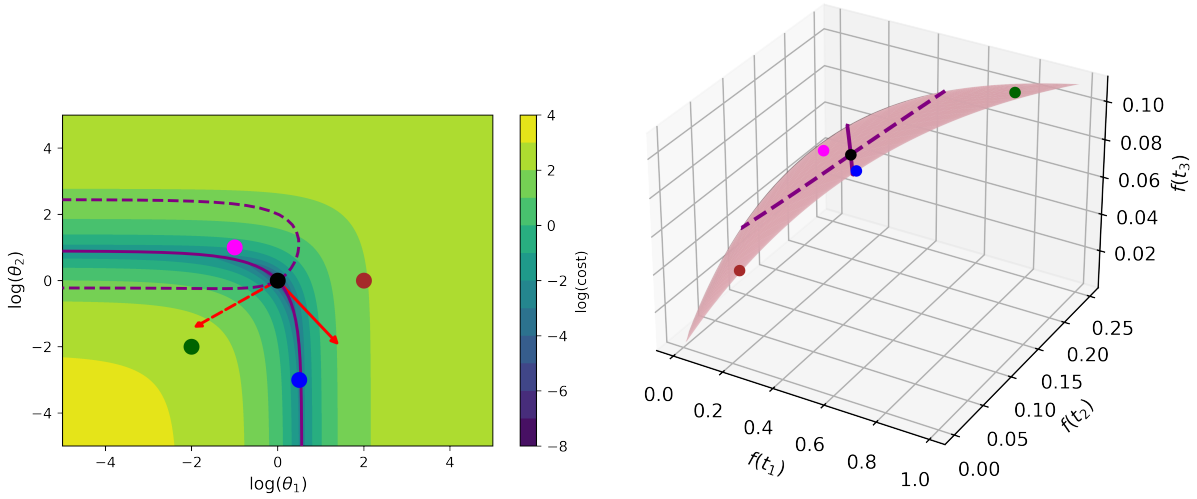
Information geometry is an approach to statistics in which a multi-parameter model is interpreted as a high-dimensional manifold. We study this manifold using computational differential geometry that allows us to extract the key geometric and topological features of the model manifold. These features shed light on issues related to UQ.

As we have seen in Sec. 2.1, a multi-parameter model makes a set of predictions, $f_m(\theta)$, that we interpret as a vector in data space. That is to say, the model is a mapping between parameter space and data space:

$$f : \mathcal{D} \subseteq \mathbb{R}^N \to \mathbb{R}^M. \tag{2.15}$$

Conceptually, the model manifold is constructed by mapping all possible parameter values to their corresponding predictions in data space, i.e., the model manifold is the image of parameter space under the model map, illustrated for the toy model (Eq. 2.4) in Fig. 2.6.

Critically, notice that the manifold is bounded by two one-dimensional segments. We focus on the boundaries of the model manifold as they are the geometric feature most relevant to parameter uncertainty. In parameter space, we have seen there can be large or infinite uncertainties when cost contours do not close, i.e., confidence regions extend to the limits of the parameter domain. Infinite, high entropy regions of parameter space are mapped to finite regions near the boundary of the model manifold, thus these contours are generic when the model manifold is bounded. Since the



(a) **(b)**

Figure 2.6 Model manifold of the toy model in Eq. 2.4. Each point in parameter space (a) corresponds to a set of predictions in data space (b). The set of all possible predictions trace out the model manifold. For reference, colored points are the same as those in Fig. 2.1. The red arrows in parameter space shows the eigenvectors of the FIM. Geodesics (solid and dashed curves) relate manifold structures, such as manifold boundaries, to parameter space.

cost corresponds to distance in data space, the images of cost contours are approximately concentric circles on the model manifold. Each segment of the boundary is a manifold of co-dimension one. Associated with each boundary segment is a parameter, or combination of parameters, that are *practically unidentifiable* at some level of statistical confidence.

We use geodesics, i.e., distance minimizing curves, on the model manifold to find the unidentifiable parameter combination associated with each boundary segment [34]. We approximate geodesics curves along the model manifold by numerically solving the geodesic equation

$$\frac{\partial^2 \theta^i}{\partial \tau^2} = -\sum_{j,k} \Gamma^i_{jk} \frac{\partial \theta^j}{\partial \tau} \frac{\partial \theta^k}{\partial \tau}, \tag{2.16}$$

where

$$\Gamma^i_{jk} = \sum_{l,m} \left(\mathcal{I}^{-1}\right)^{il} \frac{\partial y_m}{\partial \theta^l} \frac{\partial^2 y_m}{\partial \theta^j \partial \theta^k} \tag{2.17}$$

are the so-called Christoffel symbols, $\mathcal{I}$ is the FIM, θ are the parameters, and τ is the arc length of the geodesic along the model manifold. We numerically solve for the path of the geodesic by treating this equation as an initial value problem where the initial position is given by the nominal parameter values and the initial direction is given by the sloppiest eigenvector. We provide a simple example script for calculating geodesics on github [64]. To illustrate this, we again turn to the toy model in Eq. 2.4. Fig. 2.6 shows two geodesics that pass through the best fit point on the model manifold, the solid and dashed curves in both parameter space (a) and on the model manifold (b). Notice how the geodesics rotate in parameter space to naturally follow the cost contours and align with the unidentifiable parameters. As the geodesic curves approach the edge of the model manifold, we see that either $\theta_1 \to 0$ or $\theta_2 \to 0$. From the correspondence between the parameter space picture and the data space picture, we deduce that the upper boundary segment on the model manifold corresponds to $\theta_1 \to 0$ while the lower segment corresponds to $\theta_2 \to 0$. These limiting values indicate which combinations of parameters are unidentifiable and have unbounded uncertainties.

In this simple example, the two boundary segments are already aligned with parameters of the model. In more realistic models, boundary segments often correspond to the coordinated

combinations of bare parameters. In these cases, we will use geodesics to identify the correlation and find a more natural, identifiable reparameterization.

2.6 Interatomic Potentials and Tests

In this study we apply the methods described in Sec. 2.2 - 2.5 to empirical IPs taken from the OpenKIM repository [11, 65]. The OpenKIM framework has a standardized collection of models, data, and tests for computing materials properties that make the UQ process reproducible and transferable. For this study we chose the Lennard-Jones and Morse potentials for silicon and nickel, respectively, to validate methods and demonstrate general principles on low-dimensional models. We then extend the investigation to the molybdenum disulfide (MoS_2) system using the more complex Stillinger–Weber potential [2].

These potentials are categorized as cluster potentials. Given a system with N atoms, the total potential energy, $\mathcal{V}$, is

$$\mathcal{V} = \sum_{\substack{i,j=1 \\ i<j}}^{N} \phi_2(\mathbf{r}_i, \mathbf{r}_j) + \sum_{\substack{i,j,k=1 \\ i<j<k}}^{N} \phi_3(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) + \dots, \tag{2.18}$$

where ϕ_n denotes the n-body potential function and $\mathbf{r}_i$ is the position of atom i.

The Lennard-Jones (LJ) potential is a pair potential, i.e., Eq. 2.18 only consists of the two-body (pair-wise) interaction term and the the higher order potential functions are set to zero. The pair-wise interaction has two parameters, given by

$$\begin{aligned} \phi_{\mathrm{LJ}}(r_{ij}) &= 4\varepsilon \left(\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right) + \Delta, \\ \Delta &= -4\varepsilon \left(\left(\frac{\sigma}{r_{\mathrm{cut}}} \right)^{12} - \left(\frac{\sigma}{r_{\mathrm{cut}}} \right)^{6} \right), \end{aligned} \tag{2.19}$$

where $r_{ij} = \|\mathbf{r}_i - \mathbf{r}_j\|$ is the distance between atoms i and j. The potential is only non-zero when $r_{ij} < r_{\mathrm{cut}} = 7.91118$ Å. The parameter ε is an energy scaling factor in the potential, while σ is

related to the equilibrium distance of the pair interaction. The shifting factor, Δ, is chosen so that the potential is continuous at r_{cut}. The parameter values for silicon are $\varepsilon = 3.17431$ eV and $\sigma = 1.9778$ Å [66–70].

The Morse potential is also a pair potential, similar to the LJ potential. The pair-wise interaction with three parameters is given by

$$\begin{aligned} \phi_{\text{M}}(r_{ij}) &= \varepsilon \left(-e^{-2C(r_{ij}-r_0)} + 2e^{-C(r_{ij}-r_0)} \right) + \Delta, \\ \Delta &= -\varepsilon \left(-e^{-2C(r_{\text{cut}}-r_0)} + 2e^{-C(r_{\text{cut}}-r_0)} \right), \end{aligned} \tag{2.20}$$

where ε is an energy scaling factor, r_0 is the equilibrium distance and C controls the width of the potential well. Again, the potential is only non-zero when $r_{ij} < r_{\text{cut}} = 9.75476$ Å and Δ is chosen such that the potential is continuous at r_{cut}. The parameter values for nickel are $\varepsilon = -0.4205$ eV, $C = 1.4199$ Å^{-1}, and $r_0 = 2.78$ Å [71–73].

We use these pair potentials to predict the unrelaxed energy and forces of silicon (LJ potential) and nickel (Morse potential) atoms in a randomly perturbed body-centered triclinic configuration with a periodic boundary condition. The lattice parameters are given below [74, 75].

$$\begin{aligned} a &= 3.1287525 \text{ Å} & \alpha &= 87.25318054968444^\circ \\ b &= 3.15146 \text{ Å} & \beta &= 93.34074777413502^\circ \\ c &= 3.13121044 \text{ Å} & \gamma &= 91.23134462011188^\circ \end{aligned}$$

We also use a random triclinic configuration with 64 silicon atoms [76] with several other cluster potentials in a broader survey of the Fisher information. We take as data the predicted energy and forces at the default parameters and assume 10% error bars (for forces, we use 10% of the magnitude of the force vector).

Note that these structures are not the ground states for silicon or nickel, but they are tests available in OpenKIM that are convenient for validating methods on low-dimensional models. We will see that they clearly illustrate the problems that the sloppiness of the model brings to the

standard UQ methods and the phenomena that we will discuss later are generic to any atomic configuration used.

We extend the analysis to the Stillinger–Weber (SW) potential for monolayer MoS_2 [77, 78], which contains both pair-wise and three-body interactions. The two-body interaction takes the form of

$$\phi_2^{IJ}\left(r_{ij}\right) = A_{IJ}\left(B_{IJ}\left(\frac{\sigma_{IJ}}{r_{ij}}\right)^{p_{IJ}} - \left(\frac{\sigma_{IJ}}{r_{ij}}\right)^{q_{IJ}}\right)\exp\left(\frac{\sigma_{IJ}}{r_{ij} - r_{IJ}^{\text{cut}}}\right), \tag{2.21}$$

where uppercase subscripts denote the types of atoms, e.g., A_{IJ} is the parameter A corresponding to interaction between atoms of type I and type J. The three-body term is given by

$$\phi_3^{IJK}\left(r_{ij}, r_{ik}, \beta_{jik}\right) = \lambda_{JIK}\left(\cos\beta_{jik} - \cos\beta_{JIK}^0\right)^2 \exp\left(\frac{\gamma_{IJ}}{r_{ij} - r_{IJ}^{\text{cut}}} + \frac{\gamma_{IK}}{r_{ik} - r_{IK}^{\text{cut}}}\right), \tag{2.22}$$

with β_{jik} being the angle between the i–j and i–k bonds.

We calibrate this potential to predict the atomic forces in configurations near the equilibrium state, as described in [2]. Our formulation follows closely that of the original paper (e.g., we set $q_{IJ} = 0$, fix γ to be the same for all types of interaction, and use the same training set); however, we make a few changes. First, we allow parameters p_{IJ} to take any positive real value and remove the relation between σ_{IJ} and the equilibrium lattice constants of the system. We also do not require $d\phi_2/dr|_{r=d} = 0$ at the equilibrium bond length d, which removes the constraint on B_{IJ}. The remaining free parameters are A_{IJ}, B_{IJ}, p_{IJ}, and σ_{IJ} for each type of pair-wise interaction (Mo–Mo, Mo–S and S–S interactions), λ_{IJK} for S–Mo–S and Mo–S–Mo interactions, and γ.

We again choose error tolerances to be 10% of predicted values. Note that this leads to non-uniform weighting factors in our cost function, unlike reference [2]. Fitting this model leads to a new set of optimal parameter values listed in Table 2.1 for the two-body interaction term and Table 2.2 for the three-body interaction term. Other parameters that are not listed in these tables take the same values as listed in [2], such as the cutoff radii and the reference bond angle. The cost at the best fit is 1.390×10^6. Because the fitting data are forces near equilibrium, the error bars are

very small (leading to a large cost) with larger weight on configurations near equilibrium. However, in our Bayesian analysis we sample the posterior at many temperatures, effectively scaling these small error bars up to something more reasonable. The process provides a systematic study of the role of error bars in quantifying parametric uncertainty in sloppy, molecular models.

	Interaction		
Parameter	Mo–Mo	Mo–S	S–S
A (eV)	18.4310060	8.83861305	0.37463396
B	0.00641786	1.04793603	561.429270
p	4.73717813	8.26621744	2.66196913
σ(Å)	6.16940454	1.92967991	0.41904814

Table 2.1 Fitted parameters of the two-body term in the SW potential for MoS_2.

Parameter	value
$\lambda_{\text{S–Mo–S}}$ (eV)	4.28784076
$\lambda_{\text{Mo–S–Mo}}$ (eV)	14.4285026
γ (Å)	1.53800500

Table 2.2 Fitted parameters of the three-body term in the SW potential for MoS_2.

After the calibration process, we propagate the parametric uncertainty of this potential to predict the uncertainty of the change in energy as a response to the lattice stretching and compression. This calculation is done by creating MoS_2 unit cells with various in-plane lattice constants a, then relaxing the atoms in the perpendicular, out-of-plane, direction. We probe the calculation in the range $(a-a_0) \in [-0.5, 0.5]$ Å, where a_0 is the equilibrium lattice constant. Then, we compare the uncertainty to the result in [2] qualitatively.

Chapter 3

Results

Fig. 3.1 shows the eigenvalues of the FIM for the models in Sec. 2.6, evaluated at the nominal values of the parameters. Notice that they are sloppy; the eigenvalues cover many orders of magnitude, indicating that many parameters are unidentifiable from the data. To illustrate that sloppiness is a general property of IPs, we also include the eigenvalues of the FIM, evaluated at the original parameters, for the Khor–Das Sarma potential [79,80], Environment Dependent Interatomic Potential (EDIP) [81–84], and SW potential [1, 85–89] in predicting the energy and forces of the atoms in a random triclinic silicon configuration [75, 76]. Fig. 3.2 shows the participation factor [90,91] for the SW MoS_2 model, i.e. how much each parameter contributes to each eigenvector. Participation factors are calculated as the element-wise square of the eigenvectors of the FIM. We conclude that the sloppiest direction, indicated by the eigenvector with the smallest eigenvalue, is dominated by the parameter $B_{\text{S–S}}$. Similarly, we can read off the participation factors of each parameter in the other eigendirections.

Although the FIM is a local calculation, it is computationally inexpensive compared to other methods discussed here. As a result we recommend using the FIM as an initial step to UQ. We will revisit the results from the FIM when we extend the analysis and compare the results to more global methods.

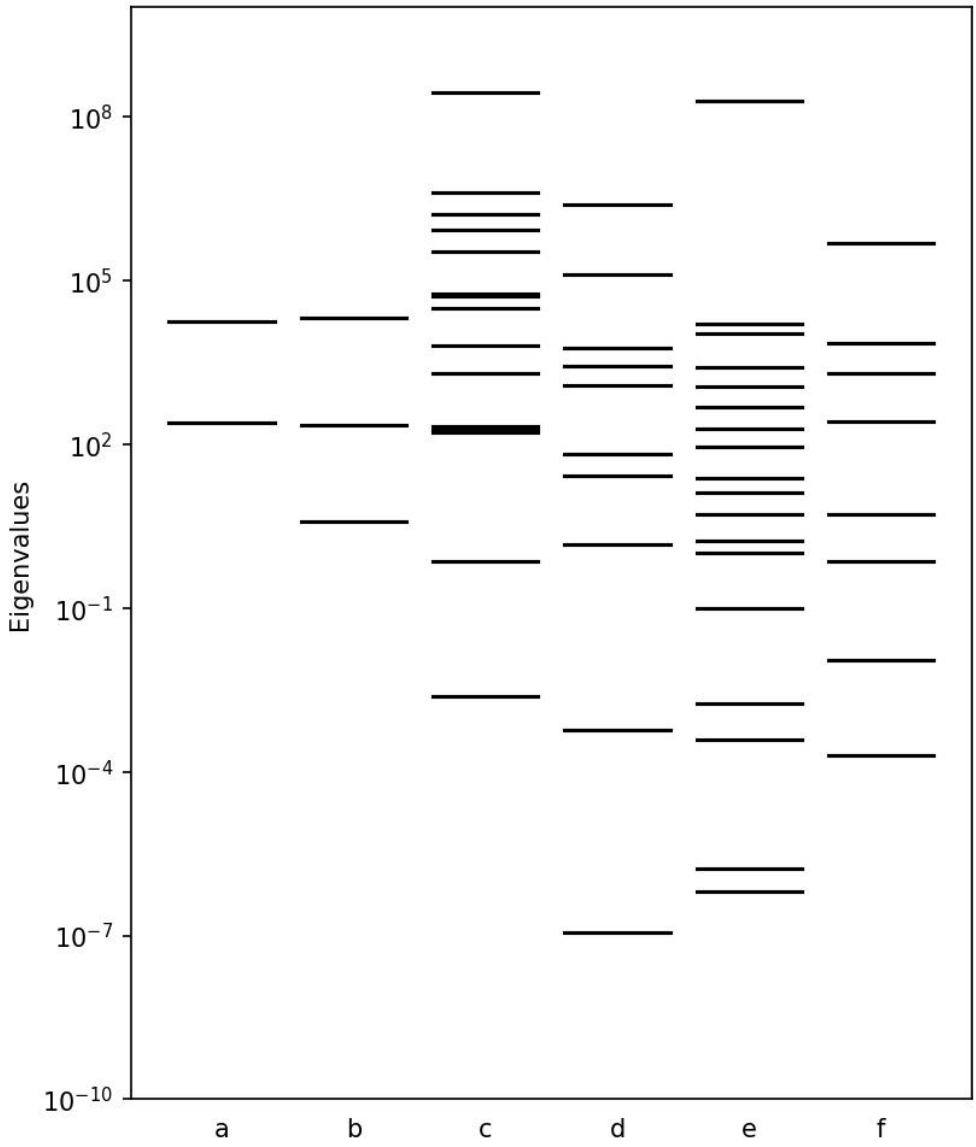


Figure 3.1 Eigenvalues of the FIM for various IPs: (a) LJ for Si, (b) Morse for Ni, (c) SW for MoS_2, and (d) Khor–Das Sarma, (e) EDIP, (f) SW, each for Si. For each model, the larger (smaller) eigenvalues represent stiff (sloppy) parameter combinations in the direction of their respective eigenvectors.

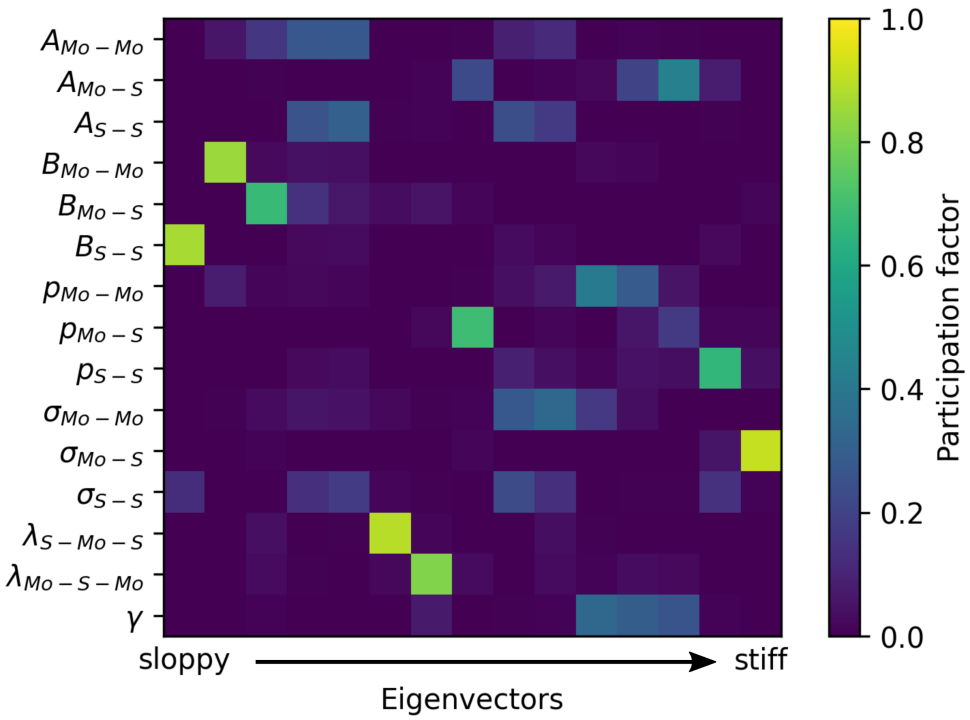


Figure 3.2 Participation factor of the SW potential for MoS_2, calculated as the element-wise square of the eigenvectors of the FIM. Each column corresponds to an eigenvector, increasing in stiffness from left to right. Participation factor shows how much each parameter contributes to each eigenvector. The parameter direction is given by the logarithm of the labels on the vertical axis. The sloppiest eigenvector (the left most column) is mostly in the $\log(B_{\text{S–S}})$ direction.

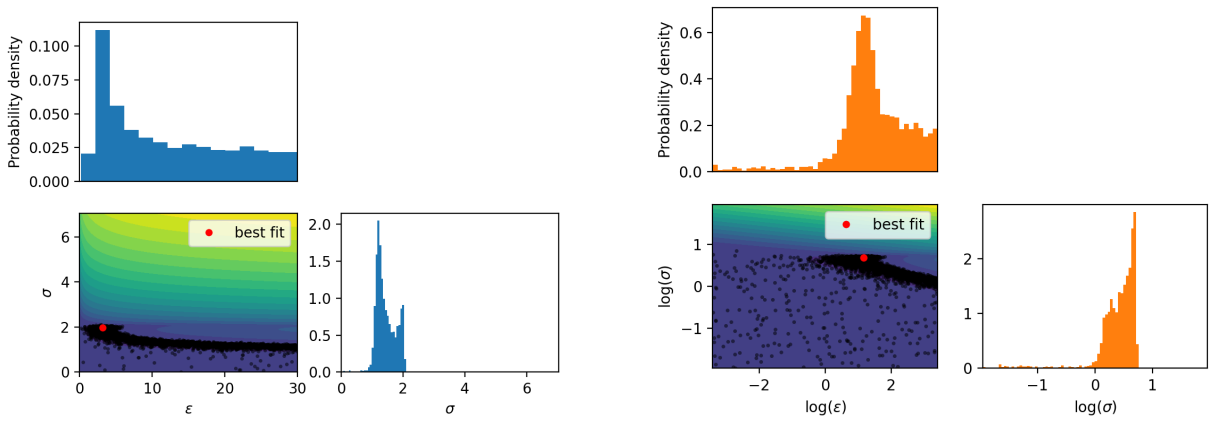


Figure 3.3 MCMC samples for the LJ potential, sampled in (a) linear and (b) log parameter scales. The original parameterization of the potential is given by the red dot. The samples are plotted against the cost contour on the lower triangle pane on each figure, with samples condensed around the low cost canyon. The marginal distributions are shown on the diagonal for each figure.

Fig. 3.3 shows the results of the Bayesian analysis for the LJ potential sampled on both the linear (Fig. 3.3a) and log scales (Fig. 3.3b). In both cases, we use a uniform prior in their respective parameter space, bounded by a rectangular region defined by $0 < \varepsilon < 30$ and $0 < \sigma < 2^{-1/6} r_{\text{cut}}$ in linear parameter space, and $|\log(\varepsilon)| < \log(30)$ and $|\log(\sigma)| < \log\left(2^{-1/6} r_{\text{cut}}\right)$ in log parameter space. The upper bound of σ is chosen so that the pair-wise equilibrium length is less than the cutoff distance. At first glance, the sampling is in good agreement with what is expected from the cost surface. Samples dominate the regions of low cost and give a visual validation that the samples are converged. Fig. 3.4 compares the marginal distributions for each parameter on both linear and log scales. Notice how the parameter scaling and, by extension, the choice of prior, can have a strong impact on the posterior distribution. On a log scale, there is a broad, flat plateau for large, negative values of $\log(\sigma)$ and $\log(\varepsilon)$. These choices affect how the uncertainties are interpreted and eventually propagated to new predictions.

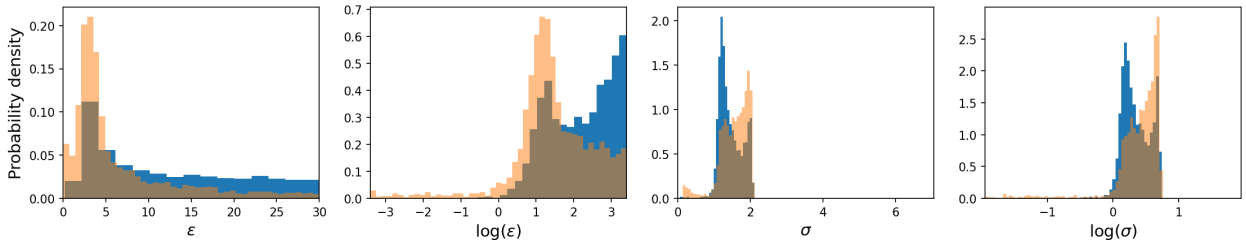


Figure 3.4 Comparison of the marginal distributions for the MCMC simulations sampled in linear (blue) and log (orange) parameter scales. A uniform prior was used in both cases. Differences in the posterior distributions reflect the role of parameter scaling and choice of prior.

In more general terms, recall that the prior defines how one measures volume in parameter space. Regions of the parameter space with large volume may dominate samples, in analogy to statistical mechanics in which high-entropy configurations can dominate an ensemble. It is also related to (though not exactly the same as) the phenomenon known as Lindley's paradox in which Bayesian and Frequentist approaches can give different results in a hypothesis test when a broad prior is used [92]. This issue can become especially subtle for sloppy models in high dimensions. These models are insensitive to coordinated changes in many parameters, indicating that there are large regions of parameter space with nearly identical fits, i.e., fits with high-entropy contributions to the posterior. In these cases, large entropic contributions may dominate their relative frequency in the posterior. The high dimensionality makes it difficult to quantify the role of the energy vs. entropy in the final sample and, by extension, justify the choice of prior.

For high-dimensional sloppy models, it is instructive to compare the results of the Bayesian and Frequentist techniques, as done in Fig. 3.5 for the LJ and Morse potentials. For the Morse potential, we also use a uniform prior, bounded by $|\log(r_0)| < \log(r_{\text{cut}})$, $|\log(C)| < \log(30)$, and $|\log(-\varepsilon)| < \log(30)$. These cases illustrate how the two methods agree in low-dimensional examples that are well-understood. Notice that the samples are very dense in regions around the paths of the profile likelihood, indicating that the posterior is energy-dominated and that there are not significant

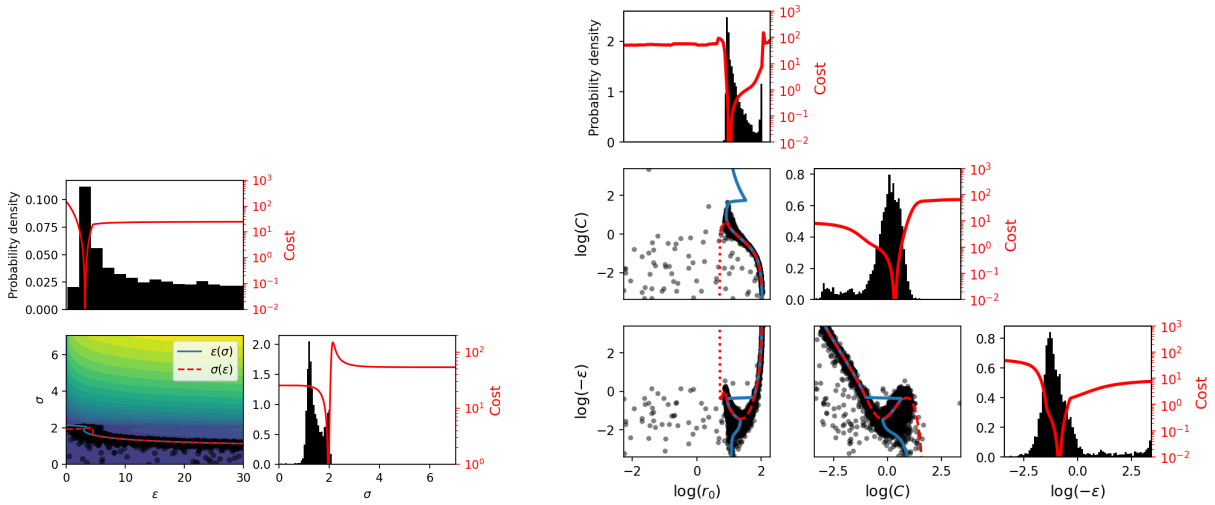


Figure 3.5 MCMC samples and profile likelihood for the (a) LJ and (b) Morse potentials. We plot the cost surface for LJ because it only contains two parameters; in general it is not possible to plot the cost surface, e.g. for Morse. On the lower triangle panes, the MCMC samples are plotted as the black points while the red and blue curves show the profile likelihood paths for the parameters on the horizontal and vertical axes, respectively. On the diagonal, we superimpose the cost profiles (red curves) on top of the marginal distribution of the MCMC samples. These plots show qualitative agreement between the two methods for low dimensional models. MCMC samples are concentrated around the profile likelihood paths, indicating that the sampling is energy-dominated and there are no significant artifacts from the choice of prior. However, there are signs of large entropy regions that could dominate the sampling at higher temperatures, e.g., evaporation at large negative values of $\log(r_0)$.

artifacts from the prior. Furthermore, the marginal distributions of each parameter are congruous with the profile likelihoods (main diagonal). However, there are hints of large-entropy regions that could become significant at higher sampling temperatures, for example, samples evaporating on the sub-optimal region at large negative values of $\log(r_0)$.

We now turn to the SW model in Fig. 3.6. We set the prior distribution to be uniform in a rectangular region, defined by $|\log(\theta_i)| < 24$, where θ_i are the parameters in this potential. Previous results have shown that a natural sampling temperature given by $T_0 = 2C_0/N \approx 1.85 \times 10^5$ (recall that $C_0 = 1.390 \times 10^6$ is the minimum cost) gives a good estimate of the systematic errors in the

model [4]. Thus, raising the temperature from $T = 1$ to $T = T_0$ transitions the sampling from the target accuracy to a more realistic estimate of the actual systematic errors. Fig. 3.6 summarizes a Bayesian sampling at four temperatures for parameters $A_{\text{S–S}}$ and $B_{\text{S–S}}$ ($T = 5.40 \times 10^{-3}\ T_0, 5.40 \times 10^{-2}\ T_0, 5.40 \times 10^{-1}\ T_0, 5.40\ T_0$). The sampling results at other sampling temperatures for the other parameters can be found in Appendix A.

At low temperatures, the profile likelihoods again agree with the Bayesian sampling. Next, we increase the sampling temperature. Recall that the temperature uniformly scales the error bars in Eq. 2.2. As the temperature rises, the uncertainty estimates in the parameters also increase; however, it does not increase uniformly in each of the parameters. At some critical temperatures, the uncertainty in a particular parameter abruptly transitions to infinity. For example, notice that, from the spread of the samples, the uncertainties in the parameters $A_{\text{S–S}}$ and $B_{\text{S–S}}$ are relatively small at $T = 5.40 \times 10^{-3}\ T_0$, but becomes effectively infinite at $T = 5.40 \times 10^{-1}\ T_0$. The posterior has transitioned from a distribution of low-temperature, energy-dominated to high-temperature, entropy-dominated samples. The higher sampling temperature has "evaporated" the parameter. We discuss this further in Ch. 4.

For the next step of UQ for this model, we propagate the parametric uncertainty and calculate the uncertainty of the change in energy as a response to lattice stretching and compression. Fig. 3.7 shows the uncertainty of this quantity of interest, calculated at several different sampling temperatures using the ensembles in Fig. 3.6. Notice that at lower sampling temperatures, such as at $T = 5.40 \times 10^{-3}\ T_0$ (blue), the uncertainties of the predicted quantities are finite. Moreover, the uncertainty in the tension domain ($a > a_0$) matches the distribution of predicted quantities from various models in [2]. However, at higher temperatures, the uncertainties diverge as the MoS_2 sheet is compressed.

At higher temperatures, some of the MCMC walkers sample regions with extreme values of parameters, near the edge of the support of the prior. These evaporated samples represent

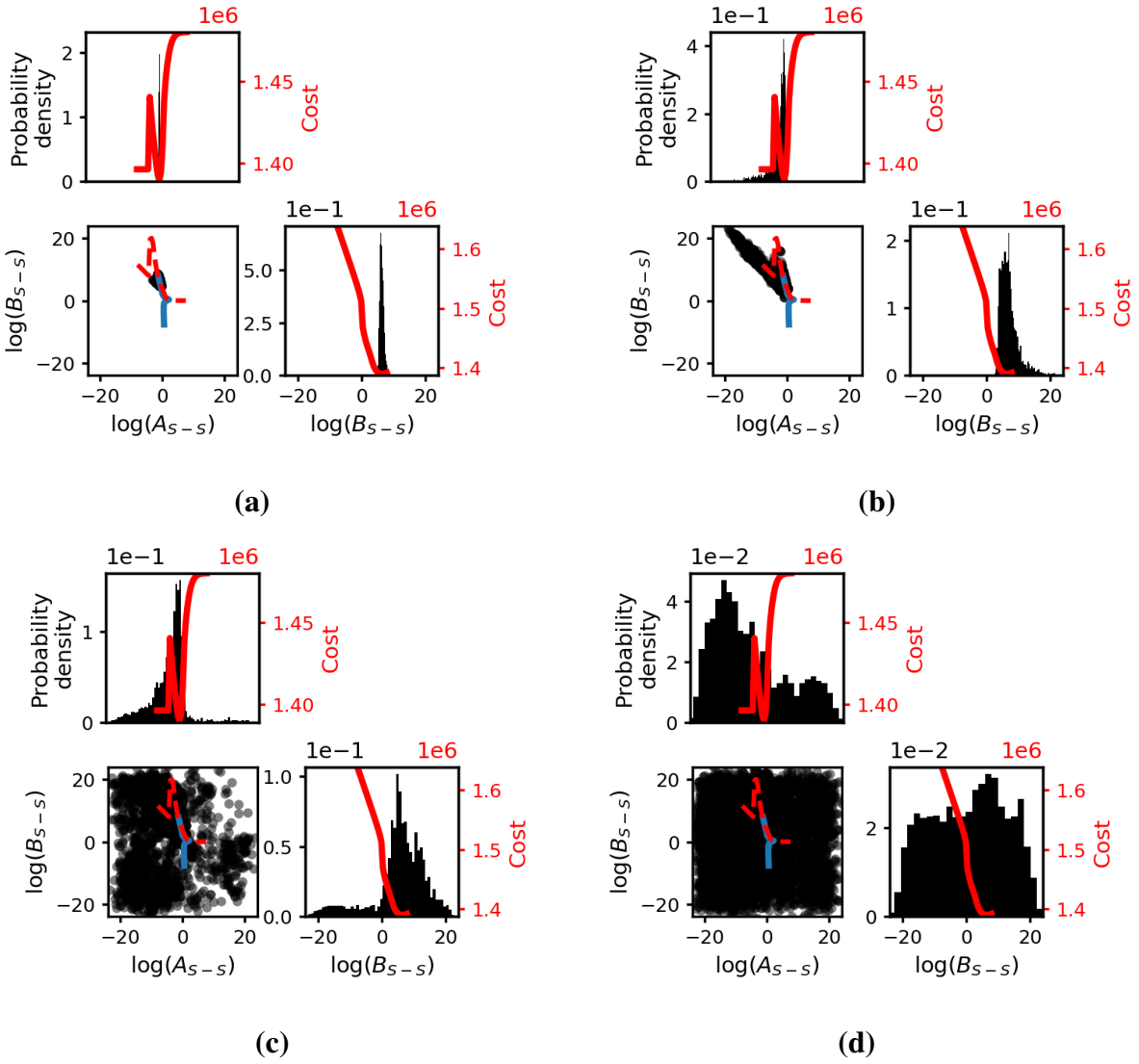


Figure 3.6 MCMC sampling for parameters A and B of the S–S interaction in the SW MoS_2 potential with a uniform prior in log parameter space at sampling temperatures: (a) $T = 5.40 \times 10^{-3}\ T_0$, (b) $T = 5.40 \times 10^{-2}\ T_0$, (c) $T = 5.40 \times 10^{-1}\ T_0$, and (d) $T = 5.40\ T_0$, where T_0 is a natural temperature. Notice that different parameter combinations evaporate at different temperature, e.g., $A_{\mathrm{S–S}}$ and $B_{\mathrm{S–S}}$ evaporate at lower temperature in a coordinated way (b), while they evaporate in all directions at higher temperature (c and d).

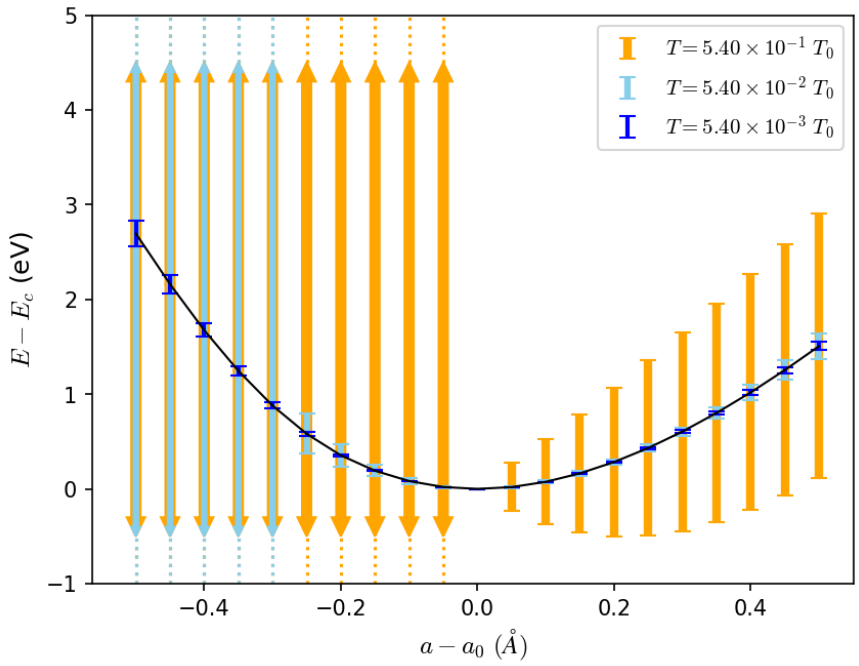


Figure 3.7 Propagated uncertainty of changes in energy as a function of lattice stretching and compression. a_0 and E_c are the equilibrium lattice constant and cohesive energy, respectively. The uncertainty of this quantity is calculated at several different temperatures from the ensembles in Fig. 3.6. Notice that the uncertainty of the energy at higher temperature diverges to infinity. This is a results of parameter evaporation, where the evaporated parameters predict infinite energy.

interactions with a very strong repulsive force in the compression domain. The magnitude of the energy grows very fast as the lattice is compressed. Consequently, the uncertainty of the energy in this domain diverges. Further, we are unable to propagate the uncertainty from the $T = 5.40\ T_0$ samples. The ensemble at this temperature contains many samples representing extreme potentials, e.g., a semi-infinite square-well potential.

The phenomenon of parameter evaporation illustrated in Fig. 3.6 has been observed previously [41]. When a parameter evaporates, its marginal posterior distribution approaches its prior. Evaporated parameters do not participate in the statistics of the model; they do not encode any information in the data and do not constrain future predictions. In other words, the effective dimen-

sionality of the model is reduced by the number of evaporated parameters. However, evaporated parameters affect statistical methods, slowing down convergence in both MCMC sampling and profile likelihood optimization. They also obscure interpretation since the evaporated parameters are often combinations of the bare parameters. UQ with these “nuisance” parameters is challenging.

Parameter evaporation is a global manifestation of the “sloppiness” phenomenon. Sloppiness was first recognized as the exponential distribution of FIM eigenvalues, as in Fig. 3.1, a local calculation. However, it was later shown using information geometry that sloppiness is a global property of the model. For sloppy models, the entire model manifold is systematically compressed into an object of low effective dimensionality, and in many practical cases, the eigenvalues of the FIM (local property) are a good estimate for the widths of the model manifold (global property) [8, 93]. We check this correspondence for the case of the SW model by comparing the eigenvalues in Fig. 3.1 with the number of effective (non-evaporated) parameters in the model at each sampling temperature. We consider a parameter to “evaporate" if the samples approach a boundary of the prior corresponding to this parameter. Although in general there is a subtle difference between the evaporated and the non-evaporated parameters, the temperature ladder we use is sparse enough that there is a clear distinction between the two. We show the comparison between the local estimate and the result deduced from MCMC in Table 3.1, and we find good agreement.

Sloppiness in high dimensions leads to cost contours that do not close and complicates the question of prior selection and the role of parameter-space entropy in the Bayesian posteriors. We now use information geometry to better understand this phenomenon, first using the LJ model as a motivating example.

Temperature (T_0)	Effective dimensionality	Local estimate
5.40×10^{-6}	15	13
1.71×10^{-6}	15	13
5.40×10^{-5}	15	13
1.71×10^{-5}	15	13
5.40×10^{-4}	15	13
1.71×10^{-4}	15	10
5.40×10^{-3}	15	10
1.71×10^{-3}	14	9
5.40×10^{-2}	11	8
1.71×10^{-2}	9	7
5.40×10^{-1}	6	5
1.71×10^{-1}	6	5
5.40×10^{0}	0	3

Table 3.1 Relation between sampling temperature, effective dimensionality, and local estimate of the effective dimensionality of the model. The effective dimensionality is the number of non-evaporated parameters at a given temperature. The local estimate is the number of eigenvalues of the Fisher information larger than a given temperature.

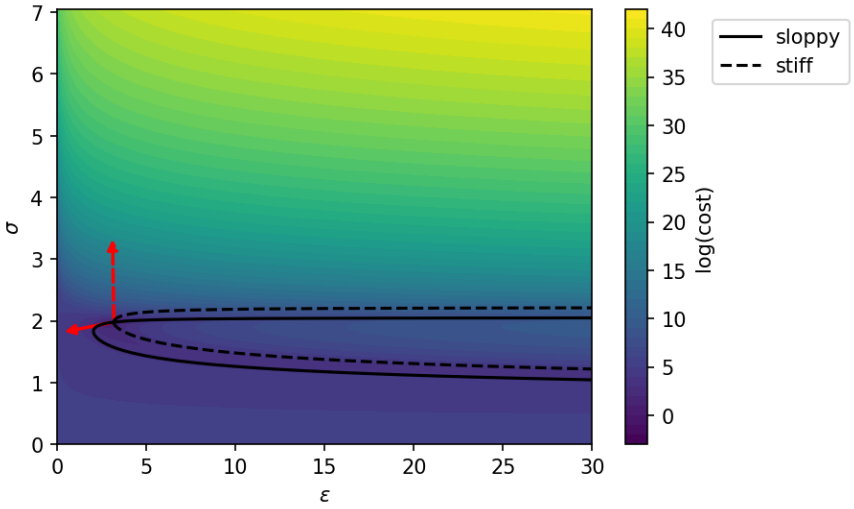


Figure 3.8 Geodesics for the LJ potential in the sloppy and stiff directions. Each geodesic starts at the best fit parameters and moves in the forward and backward directions of the sloppy and stiff eigenvectors of the FIM, shown by the corresponding arrows. By considering the difference in the scaling of the parameter axes, the eigenvectors are orthogonal to each other.

Fig. 3.8 shows four geodesics paths in parameter space radiating from the best fit. Although difficult to discern from the figure, the geodesics that found boundaries took the same asymptotic form: $\log(\varepsilon) \to \infty$, $\log(\sigma) \to -\infty$ with $\log(\varepsilon)$ diverging six times faster than $\log(\sigma)$. The significance of this result is more apparent when expressed in the so-called AB form with $A = 4\varepsilon\sigma^{12}$ and $B = 4\varepsilon\sigma^{6}$. As the original parameters, ε and σ, approach the boundary, they are correlated such that $A \to 0$ while B remains finite. This suggests the AB parameterization is a more natural parameterization, which from Eq. 2.19 gives,

$$\begin{aligned}
\phi_{\mathrm{LJ}}(r_{ij}) &= 4\varepsilon\left(\sigma^{12}\left(\frac{1}{r_{ij}^{12}} - \frac{1}{r_{\mathrm{cut}}^{12}}\right) - \sigma^{6}\left(\frac{1}{r_{ij}^{6}} - \frac{1}{r_{\mathrm{cut}}^{6}}\right)\right) \\
&= A\left(\frac{1}{r_{ij}^{12}} - \frac{1}{r_{\mathrm{cut}}^{12}}\right) - B\left(\frac{1}{r_{ij}^{6}} - \frac{1}{r_{\mathrm{cut}}^{6}}\right) \\
&\to -B\left(\frac{1}{r_{ij}^{6}} - \frac{1}{r_{\mathrm{cut}}^{6}}\right).
\end{aligned} \tag{3.1}$$

For this fitting problem, B is the identifiable parameter combination. There is a natural limit that removes the unidentifiable parameter, $A \to 0$, that leads to a physically interpretable reduced model, i.e., a purely attractive potential. Although this reduced model loses the physics of the repulsive part of the potential, the data to which the model was fit included atomic configurations that only probed the attractive regime. Thus, the geometry (1) reflects the information content of the data, (2) explains the correlations among the inferred parameters, (3) isolates unidentifiable combinations of parameters, and (4) suggests reduced models for simplifying the statistics.

An analogous calculation on the Morse potential reveals many of the same themes. We find the geodesic initially aligned with the sloppiest eigenvector of the FIM, eventually approaches a boundary in which $C \to 0$ and $\varepsilon \to -\infty$ as seen in Fig. 3.9. Notice that this geodesic aligns with the MCMC results, i.e., the low-cost canyon. As with LJ, this geodesic suggests a natural reparameterization of the model:

$$k = -2\varepsilon C^{2}. \tag{3.2}$$

As ε and C approach extreme values the specific combination k remains finite. In this parameterization, k and r_0 are the identifiable parameter combinations, while C is unidentifiable since its uncertainty extends to $C \to 0$. Evaluating the limit $C \to 0$ at constant k and r_0 leads to the simplified model:

$$\widetilde{\phi}_{\mathrm{M}}(r_{ij}) = \frac{1}{2}k(r_{ij} - r_0)^2 - \frac{1}{2}k(r_{\mathrm{cut}} - r_0)^2. \tag{3.3}$$

In this limit, the model is a simple, harmonic potential parameterized by an equilibrium position, r_0, and a stiffness, k. This indicates the configurations do not carry enough information about the potential's anharmonicity to constrain those parameter combinations. Once again, the geometry reflects the information content of the data, explains observed correlations, isolates the unidentifiable combinations, and suggests alternative parameterizations and simplified models.

We now consider the SW potential. As before, we calculate a geodesic in the sloppiest direction and find that it encounters the boundary $A_{\mathrm{S\text{–}S}} \to 0$ and $B_{\mathrm{S\text{–}S}} \to \infty$ as shown in Fig. 3.10. Again, we reparameterize the model

$$\theta = A_{\mathrm{S\text{–}S}} B_{\mathrm{S\text{–}S}}. \tag{3.4}$$

Both $A_{\mathrm{S\text{–}S}}$ and $B_{\mathrm{S\text{–}S}}$ are unidentifiable parameters, but there is an identifiable combination given by $\theta = A_{\mathrm{S\text{–}S}} B_{\mathrm{S\text{–}S}}$. Notice that as $A_{\mathrm{S\text{–}S}} \to 0$ at constant θ, $B_{\mathrm{S\text{–}S}} \to \infty$, consistent with the evaporation in Fig. 3.10. Furthermore, considering $A_{\mathrm{S\text{–}S}}, 1/B_{\mathrm{S\text{–}S}} \to 0$ at constant θ leads to the reduced form:

$$\widetilde{\phi}_2^{\mathrm{S\text{–}S}}(r_{ij}) = \theta \left(\frac{\sigma}{r_{ij}}\right)^p \exp\left(\frac{\sigma}{r_{ij} - r^{\mathrm{cut}}}\right). \tag{3.5}$$

Fig. 3.11 shows the plots of the two-body S–S interaction term for both the original and reduced models. In addition to having fewer identifiable parameters, reduced models include the physics that is informed by the data. In this case, details about the repulsive core were removed at the boundary, resulting in a stronger repulsion at short distances that were not constrained by the fit.

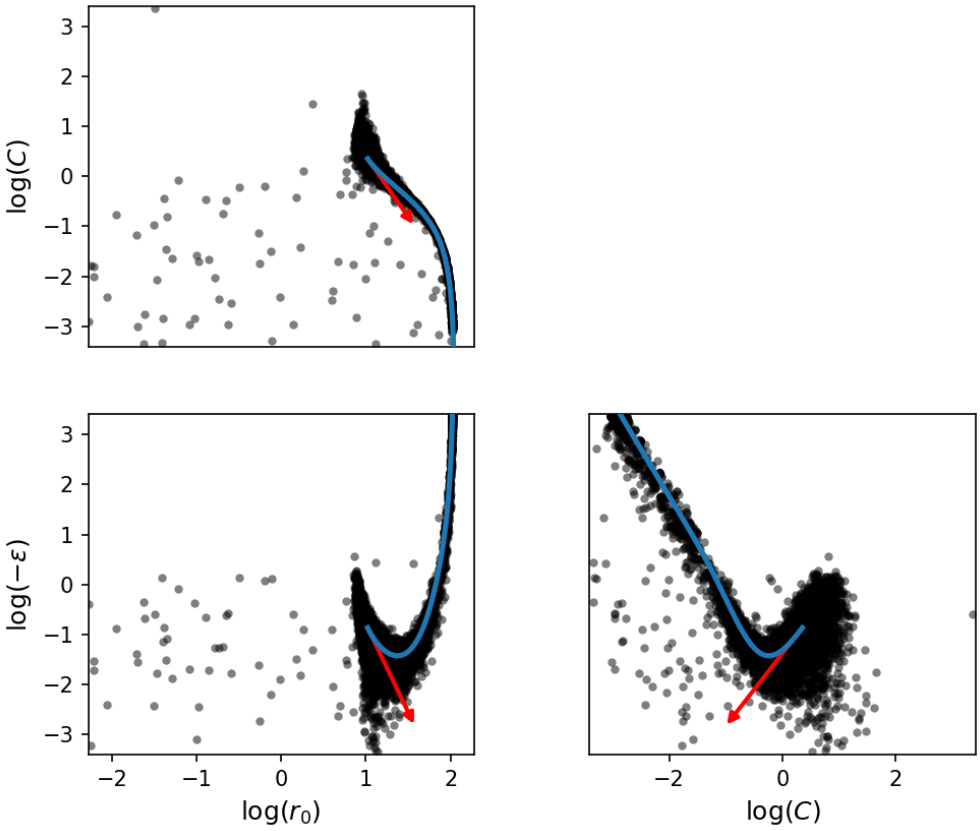


Figure 3.9 Geodesic in the sloppiest direction for the Morse potential. Geodesics (blue curves) are shown with the MCMC samples (black points) to illustrate that the geodesics follow the same low cost canyons as MCMC. These geodesics reveal specific parameter limits leading to boundaries of the model manifold, e.g., $C \to 0$ as $\varepsilon \to -\infty$.

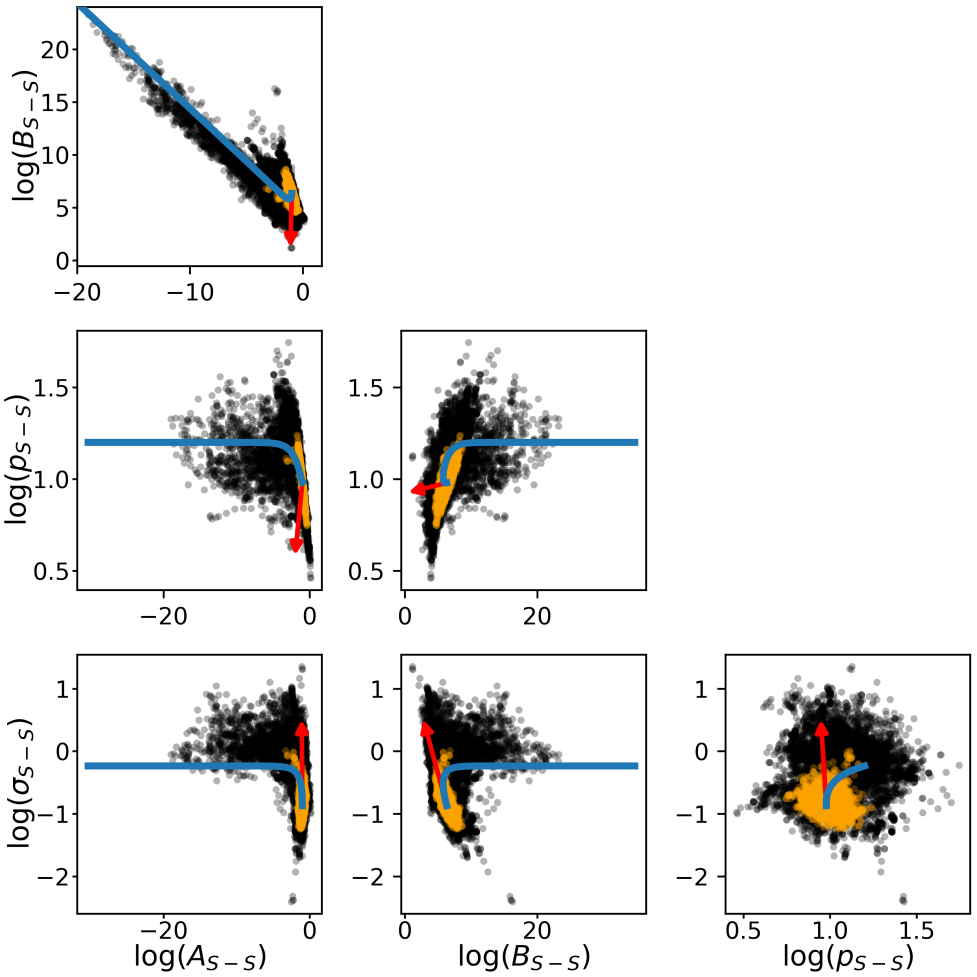


Figure 3.10 Geodesics in the sloppiest direction for the SW model. Geodesics (blue curves) are plotted with MCMC results at sampling temperatures of $5.40 \times 10^{-3}\ T_0$ (orange points) and $7.19 \times 10^{-2}\ T_0$ (black points). Axis scaling is set to show detail, not to reflect the boundaries of the uniform prior. The geodesic started in the local sloppy direction defined by the FIM, $B \to 0$ as $\sigma_{\text{S–S}} \to \infty$. Eventually, the geodesic turned to find the boundary given by the limit $A_{\text{S–S}} \to 0$ as $B_{\text{S–S}} \to \infty$.

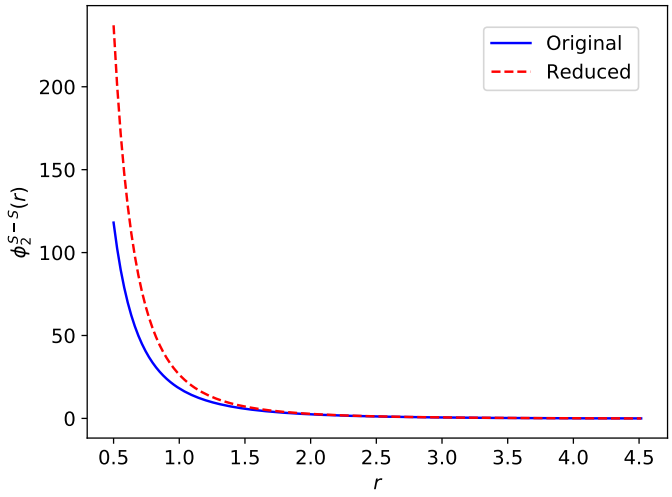


Figure 3.11 Reduced models on the boundaries have fewer identifiable parameters and abstract away irrelevant aspects of the physics. Geodesics identified the boundary defined by the coordinated limit $A_{\text{S–S}} \to 0$, $B_{\text{S–S}} \to \infty$. This figure compares the forms of the original and reduced two-body SW potentials for the S–S interaction. This reduction abstracts away details about the repulsive core of the potential.

It is interesting to compare the geodesic to the purely local analysis of the Fisher Information. The initial direction of the geodesic is given by the sloppiest eigendirection of the Fisher Information. That eigendirection, as can be seen in Fig. 3.2, is dominated by the parameters $B_{\text{S–S}}$ and $\sigma_{\text{S–S}}$. Initially, the geodesic decreases $B_{\text{S–S}}$ and increases $\sigma_{\text{S–S}}$; however, as $\sigma_{\text{S–S}}$ increases, it plays a more important role in the model. It is no longer part of the sloppy combination, so the geodesic rotates to align with $A_{\text{S–S}}$ and $B_{\text{S–S}}$, the parameters that eventually participate in the boundary (see Fig. 3.10). This simple comparison illustrates how the geodesic naturally extends the local analysis.

Chapter 4

Discussion and Conclusion

Sloppy models are often identified by their characteristic FIM spectra with eigenvalues spanning many orders of magnitude [8, 9, 32, 33]. Previous work has noted sloppiness in many contexts [9, 31, 35–39], including IPs [3, 4, 42, 42], and our results corroborate this conclusion (see Fig. 3.1). Subsequent work using information geometry showed sloppiness to be a consequence of global properties of the model, specifically that the model manifold is bounded with a hierarchy of widths [8]. In this work, we have extended the local sloppy-model analysis using (Bayesian) MCMC and (Frequentist) profile likelihoods. Each of these methods gives a unique perspective on the "global sloppiness" of the model. For example, MCMC samples evaporate sloppy parameters and cost profiles have flat plateaus resulting in diverging uncertainties. We connect these traditional statistical tools to sloppy model analysis using information geometry and geodesics. We show that the problems associated with both of these methods are features of the same underlying phenomena, sloppiness in the form of bounded model manifolds. We now discuss each of these observations in more detail.

Fig. 3.3 illustrates that cost contours of sloppy models have broad plateaus in parameter space. These regions can be thought of as high entropy states, i.e., large volumes of parameter space with approximately equal cost. Although these regions may have sub-optimal cost, they can dominate

the Bayesian posterior because of their high entropy. Previous work in sloppy models has noted this "parameter evaporation" when the posterior becomes dominated by sub-optimal, high-entropy samples [41]. Geometrically, the model maps these high-entropy regions to compressed areas near the boundaries of the model manifold. Parameter evaporation is thus the Bayesian manifestation of sloppiness that is a consequence of model boundaries.

Model boundaries have different distances from the data. This implies that the height of the cost plateau varies in each parameter direction. Consequently, specific parameters evaporate at different sampling temperatures. This is analogous to particles in a classical finite potential well; only those particles (MCMC walkers) with high enough energy can escape the well (cost surface). Fig. 3.6 shows that different parameter combinations evaporate at different temperatures. Indeed, previous studies in IPs have lowered the sampling temperature specifically to avoid parameter evaporation [3, 4]. However, for a sufficiently broad prior, parameters evaporate at any non-zero temperature, although the evaporation time may be very long for large cost barriers, making it difficult to assess convergence.

Table 3.1 reports the number of evaporated parameters for different sampling temperatures for our SW potential. Note that the number of identifiable parameter combinations, i.e., non-evaporated at a given temperature, correlates with the number of eigenvalues above that temperature. Elsewhere, it has been shown that the eigenvalues of the FIM are a good approximation for the widths of the model manifold [8, 93]. Since sampling temperature corresponds to a distance in data space, Table 3.1 confirms that the eigenvalues are a reasonable (local) approximation for the widths of the model manifold in IPs.

Most of the challenges associated with formulating and performing a Bayesian analysis of a sloppy IP result from the interplay of entropy and energy in the posterior. Recall that the Bayesian prior acts as a measure on the parameter space (see Sec. 2.4), i.e., it quantifies the entropy associated with volumes of parameter space. The ambiguity in the choice of prior leads to the issues we report

here. We advocate comparing MCMC with a Frequentist method to assess the effect the prior has on the posterior, as we have done using profile likelihoods.

The profile likelihood analysis, being a Frequentist method, does not use a prior. As such, global sloppiness manifests itself differently. The cost profiles exhibit plateaus that asymptotically approach constant values, as seen in Figs. 3.5 and 3.6. Uncertainty in a parameter is set by selecting a level of statistical significance, e.g., 95% confidence interval, and identifying those parameter values with cost less than the corresponding cost threshold. As the cost threshold approaches that of the plateau, the uncertainty diverges. This leads to uncertainty metrics that are very sensitive to the level of statistical significance, making it difficult to draw conclusions from the UQ analysis.

Another complication due to sloppiness is related to parameter correlations. Sloppy canyons and plateaus do not naturally align with the parameter axes due to correlations in the parameters. These correlations can bend sloppy canyons, as seen in Fig. 3.5a for LJ. As the profile likelihood projects bending canyons onto parameter axes, correlation is lost and the results are misleading. Using a more natural parameterization, motivated by information geometry, weakens parameter correlations and unwinds the canyons asymptotically aligning them with the parameter axes.

Geodesics extend the local FIM analysis to a global regime. For example, consider Fig. 3.10. The geodesic initially pointed in the sloppiest direction, as indicated by the FIM, but changed directions to follow the global sloppiness as it approached the manifold boundary. This behavior is due to non-linearity in the model, and is known as parameter-effects curvature [94].

The global nature of geodesics is used to find boundaries of the model manifold, revealing the cost plateaus and suggesting natural parameterizations of the model. Again, consider Fig. 3.10 where the geodesics found the manifold boundary represented by the parameter limits $A_{\text{S–S}} \to 0$ and $B_{\text{S–S}} \to \infty$. This limit demonstrates a more natural parameterization of the model with parameters $\varepsilon = 1/B_{\text{S–S}}$ and $\theta = A_{\text{S–S}}B_{\text{S–S}}$, where ε is strictly non-negative. In this parameterization, only one parameter participates in the boundary, i.e., $\varepsilon \to 0$ while $\theta = \mathcal{O}(1)$. The sloppy direction aligns

with a single parameter ε that is (mostly) uncorrelated from θ. Furthermore, rather than a diverging confidence interval for $B_{\text{S–S}}$, the confidence intervals for ε and θ remain finite.

This new parameterization suggests a simplified model in which ε has been removed. The Manifold Boundary Approximation Method (MBAM) is a tool that utilizes information geometry to find these boundary limits and construct reduced models. In this specific example, the reduced model would correspond to $\varepsilon \to 0$ while holding $\theta \sim \mathcal{O}(1)$. This limit removes the sloppy parameter from the model, leaving the identifiable combination, $\theta = A_{\text{S–S}} B_{\text{S–S}}$. Performing MBAM before UQ, i.e., finding less sloppy models, would remove the challenges we have discussed. In this paper, we have performed the first step of MBAM by using geodesics to find manifold boundaries.

Reduced models often do not transfer well to new predictive regimes. However, they make new predictions with higher levels of certainty. When the large parametric uncertainties of sloppy models are propagated to new predictions, the resulting uncertainties can be large or infinite. Reducing the sloppiness of models decreases parametric uncertainty as well as the propagated uncertainty in new predictive regimes. Future work will continue this process and perform UQ on reduced models.

We have shown that sloppy models lead to ill-posed UQ problems. For Bayesians, the challenge is how to unambiguously select a prior that does not lead to large-entropy contributions in the posterior. For Frequentists, the challenge is sensitivity to the confidence level and plateaus that do not naturally align with the bare parameters (i.e., occur due to correlations among parameters). By identifying the root cause of these problems, we hope this work will lead to more transparency in the future UQ studies for IPs. In particular, information geometry suggests solutions to these issues by identifying natural parameterizations near boundaries that provides simplified, less-sloppy models.

In conclusion, we provide suggestions both for model developers and UQ practitioners alike. For developers of empirical potentials, we recommend using the FIM to assess how parameters locally affect calculated quantities. To extend this analysis to a global regime, we recommend

using geodesics to identify more natural parameterizations as well as additional training data that is needed to identify model parameters. For example, geodesic calculations for the LJ potential above identified that the training data only contained information about the attractive part of the potential. To fully identify the LJ parameters additional data that probe the repulsive regime is needed. Using these geodesics to reduce the model with an MBAM step will decrease model sloppiness and improve future UQ. This process can then be iterated starting from the reduced model until a simple, yet accurate model is attained.

For performing UQ of IPs, we recommend starting with the FIM analysis to assess the sloppiness of the model. This analysis also provides a local estimate of which parameters evaporate at a given sampling temperature. If performing UQ with MCMC, we recommend using several different sampling temperatures, including the natural temperature [7], and some alternative priors. Then, we advocate comparing the sampling results to geodesics, a frequentist method, to assess the effect of the Bayesian prior on parameter uncertainty. Additionally, researchers can perform other frequentist analysis, e.g., profile likelihood. Finally for an extended UQ analysis, researchers can perform model reduction, in which case we recommend the iterating the steps in the previous paragraph.

Appendix A

UQ Results for SW Model

In the following figures, we present the profile likelihood results and the MCMC samples for all parameters of the SW model. This model predicts the atomic forces in an MoS_2 monolayer at several sampling temperatures. The sampling temperatures are given with respect to the natural temperature $T_0 \approx 1.85 \times 10^5$. On the lower triangle panes, the samples are plotted as the black points while the red and blue curves show the profile likelihood paths for the parameters on the horizontal and vertical axes, respectively. On the diagonal, we superimpose the cost profiles (red curves) on top of the marginal distribution of the MCMC samples.

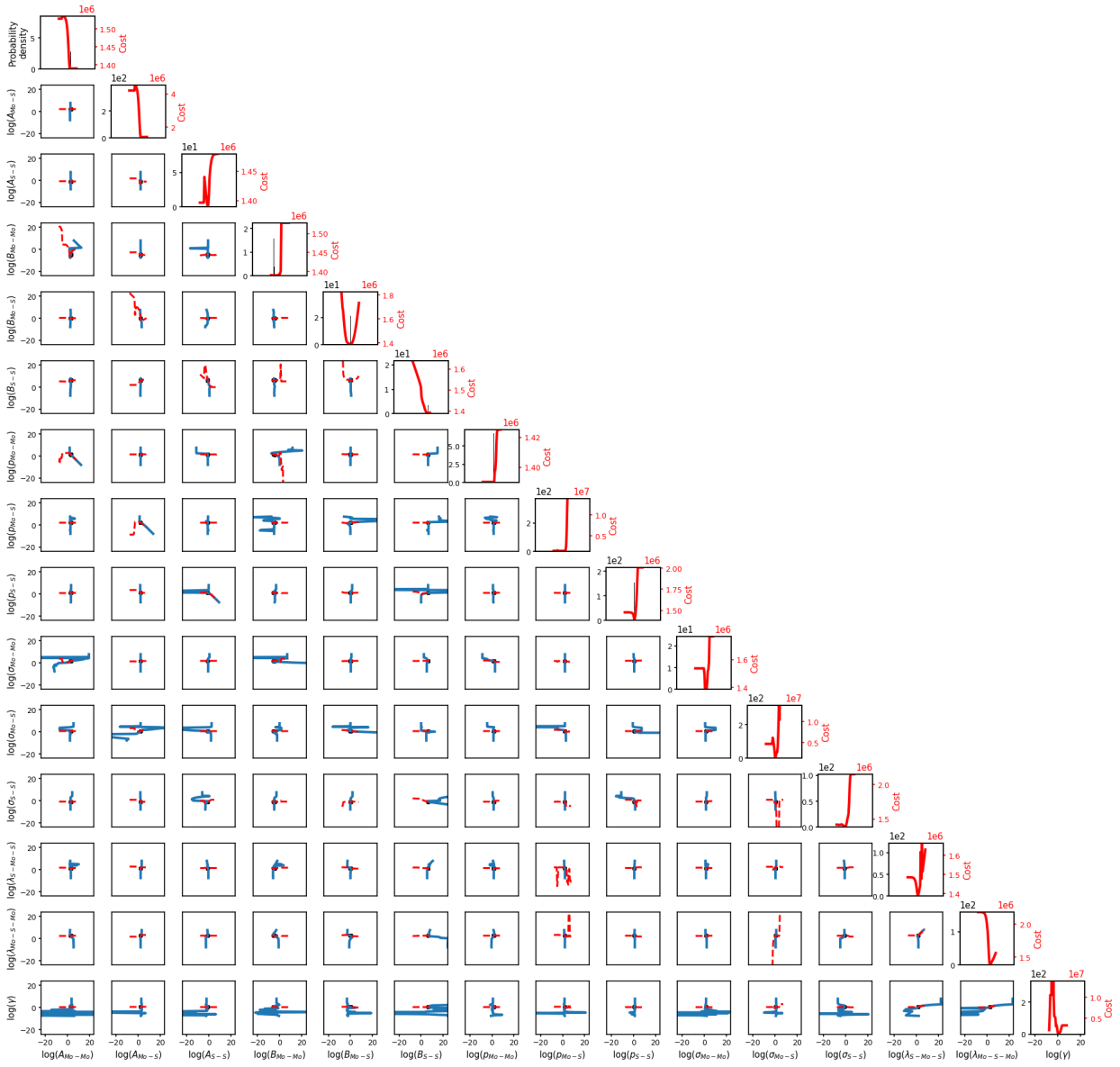


Figure A.1 Sampling Temperature 5.40×10^{-6} T_0

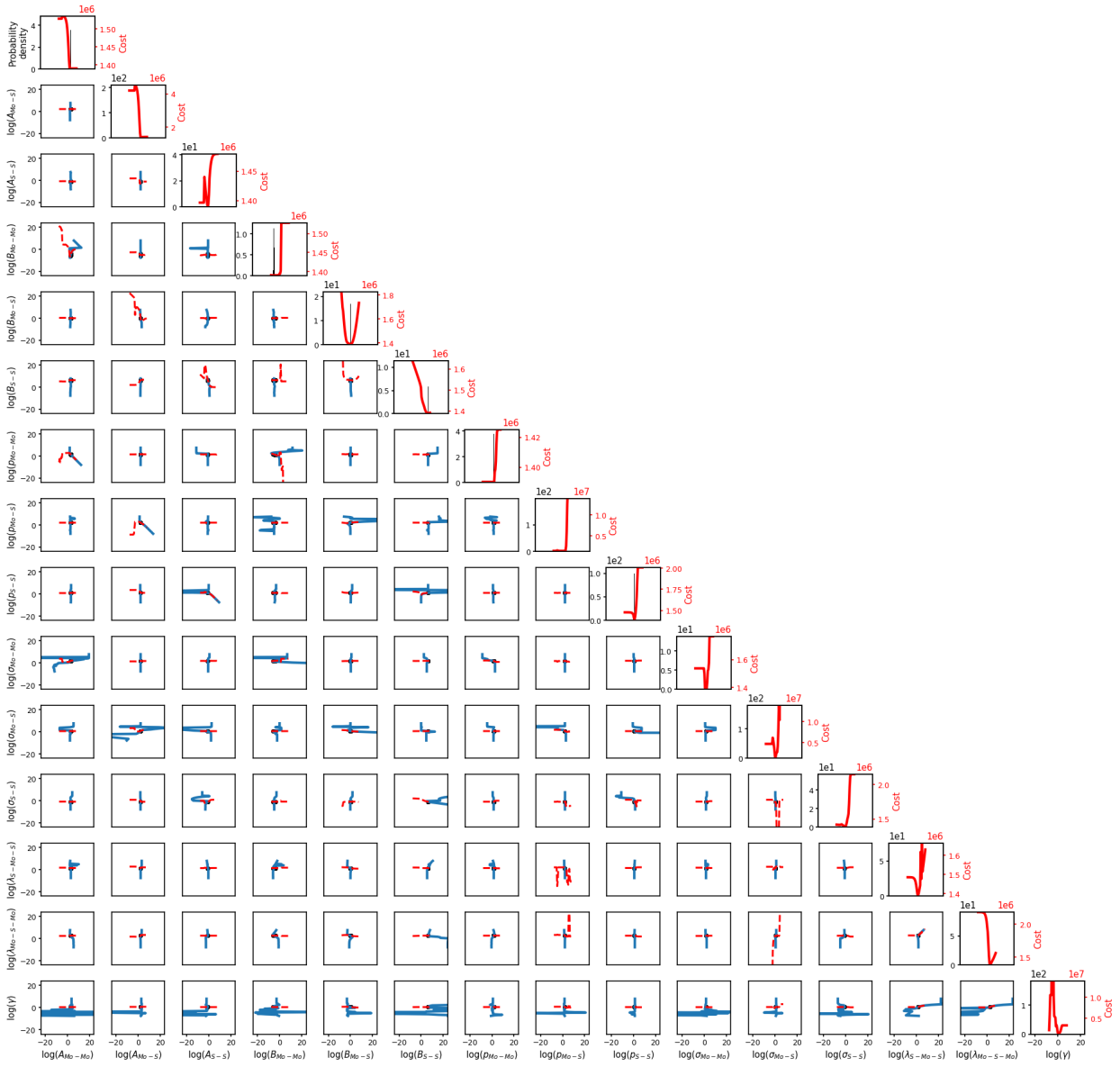


Figure A.2 Sampling Temperature 1.71×10^{-6} T_0

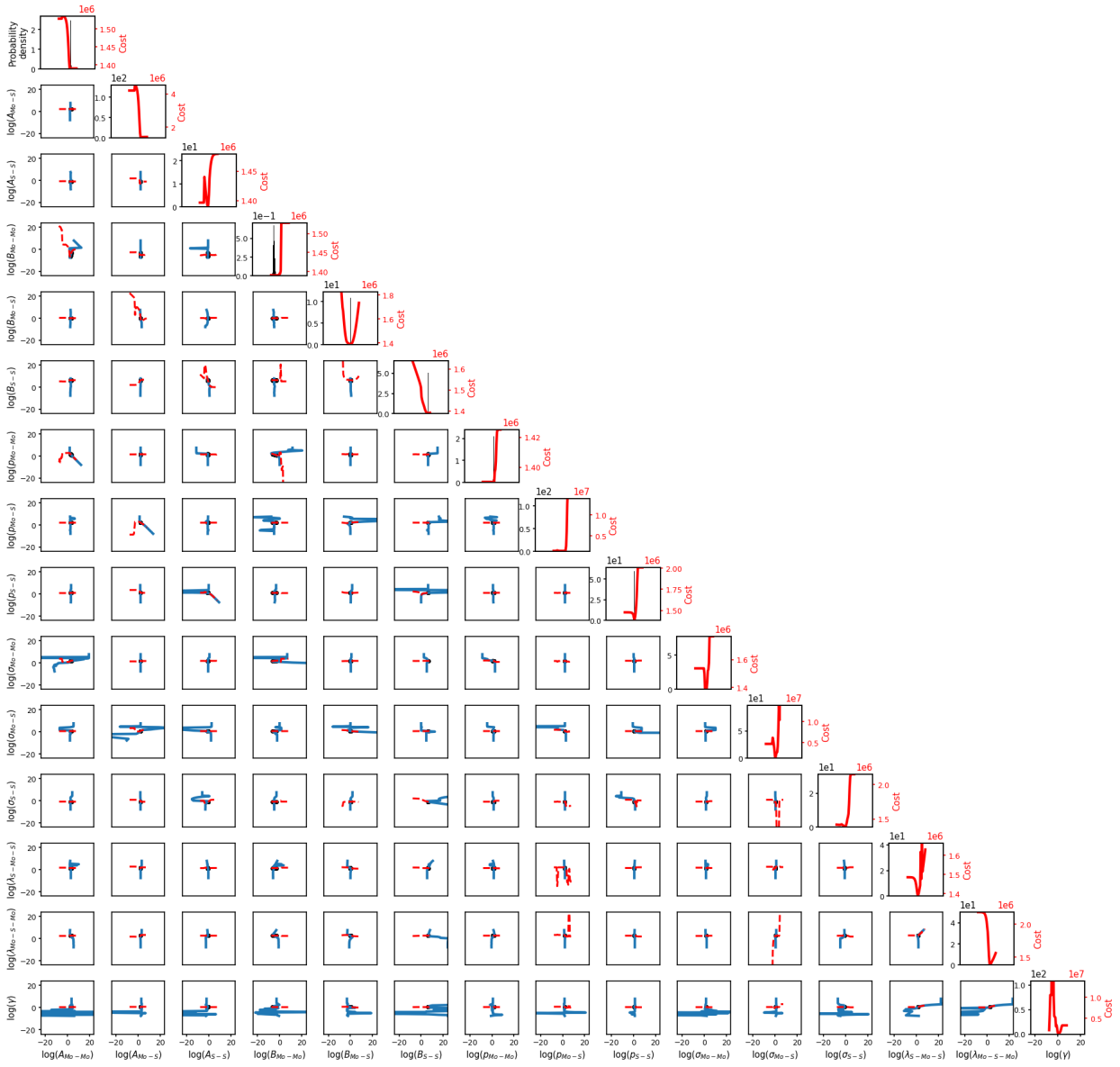


Figure A.3 Sampling Temperature 5.40×10^{-5} T_0

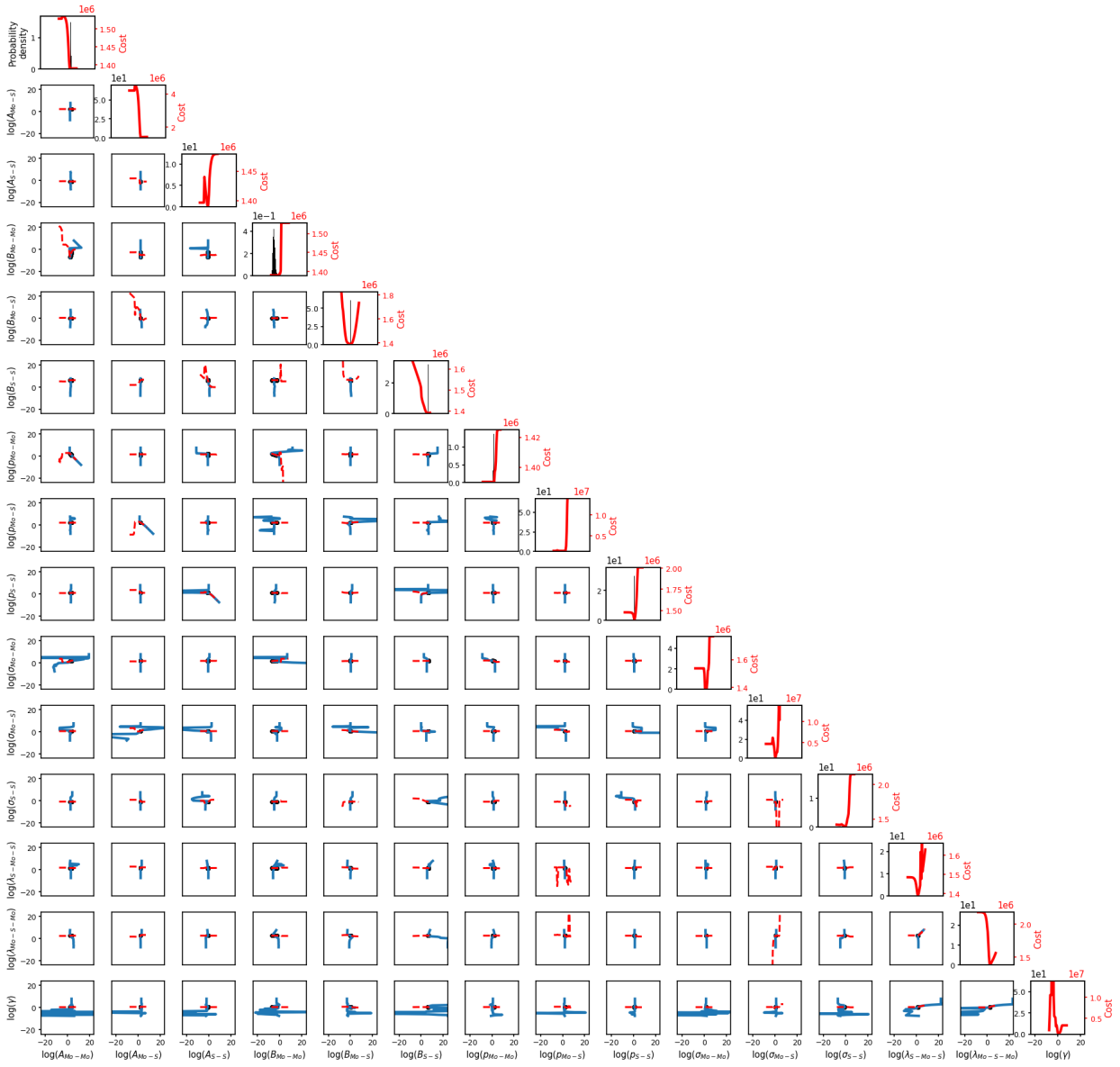


Figure A.4 Sampling Temperature 1.71×10^{-5} T_0

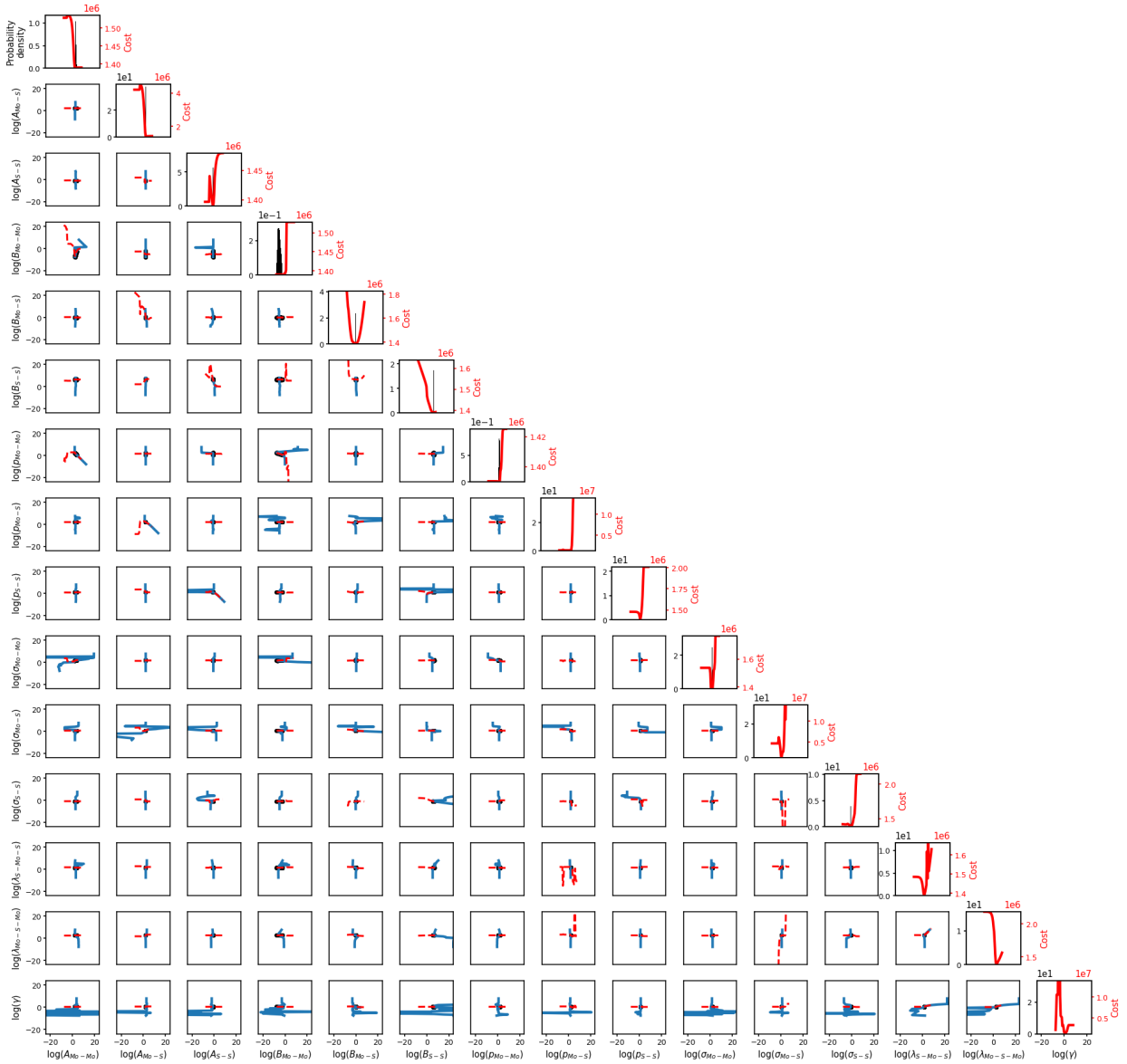


Figure A.5 Sampling Temperature 5.40×10^{-4} T_0

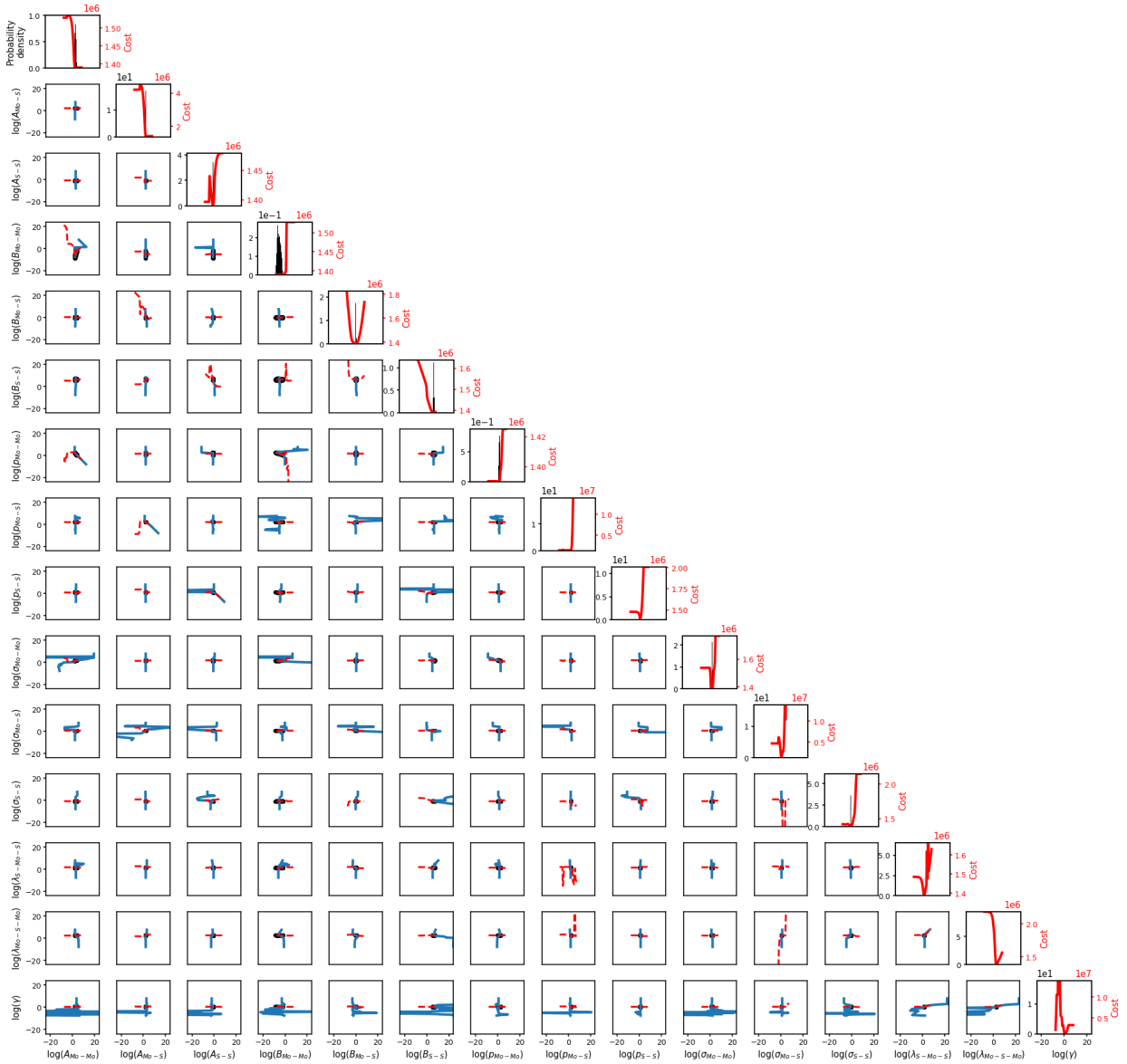


Figure A.6 Sampling Temperature 1.71×10^{-4} T_0

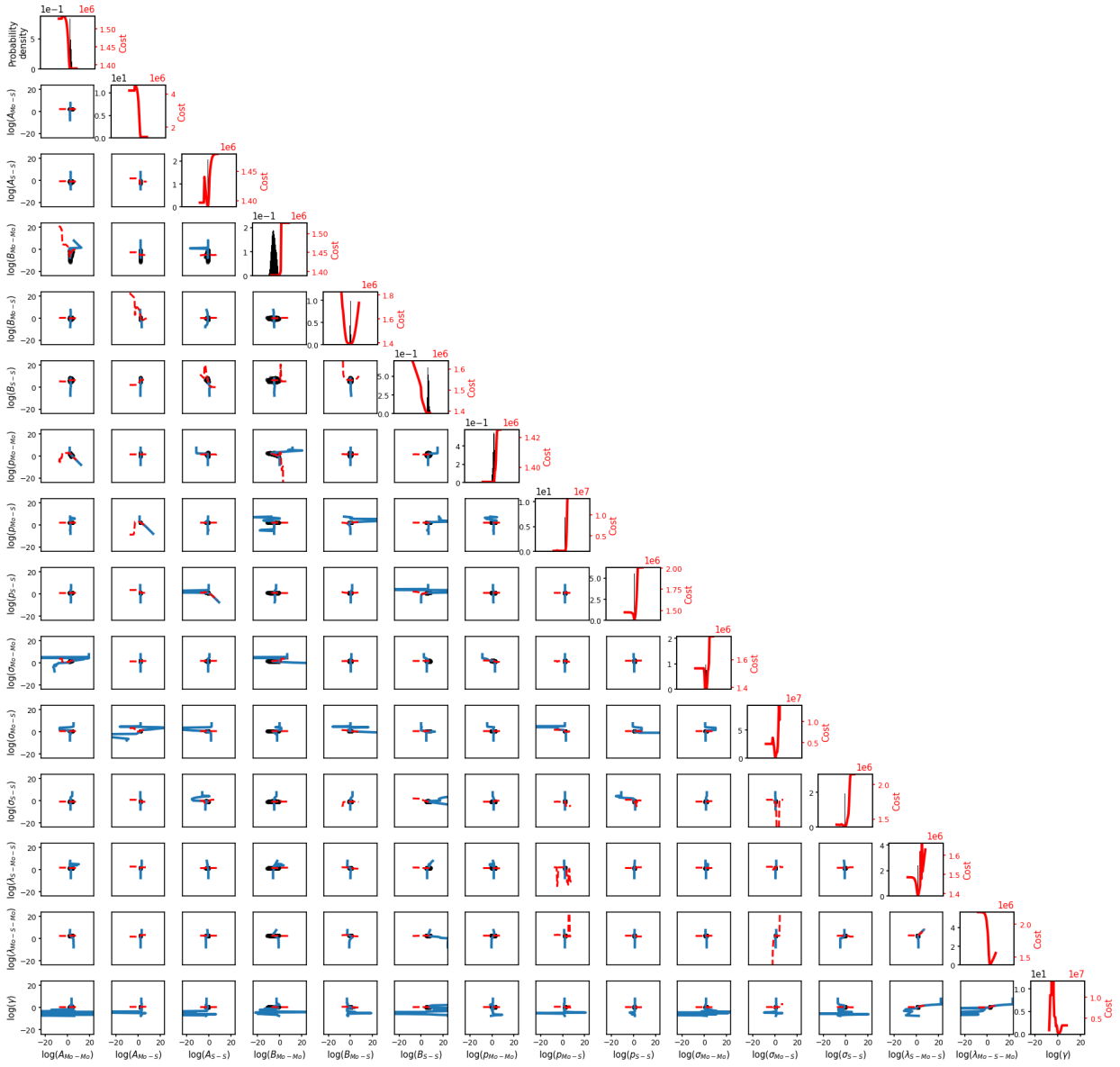


Figure A.7 Sampling Temperature 5.40×10^{-3} T_0

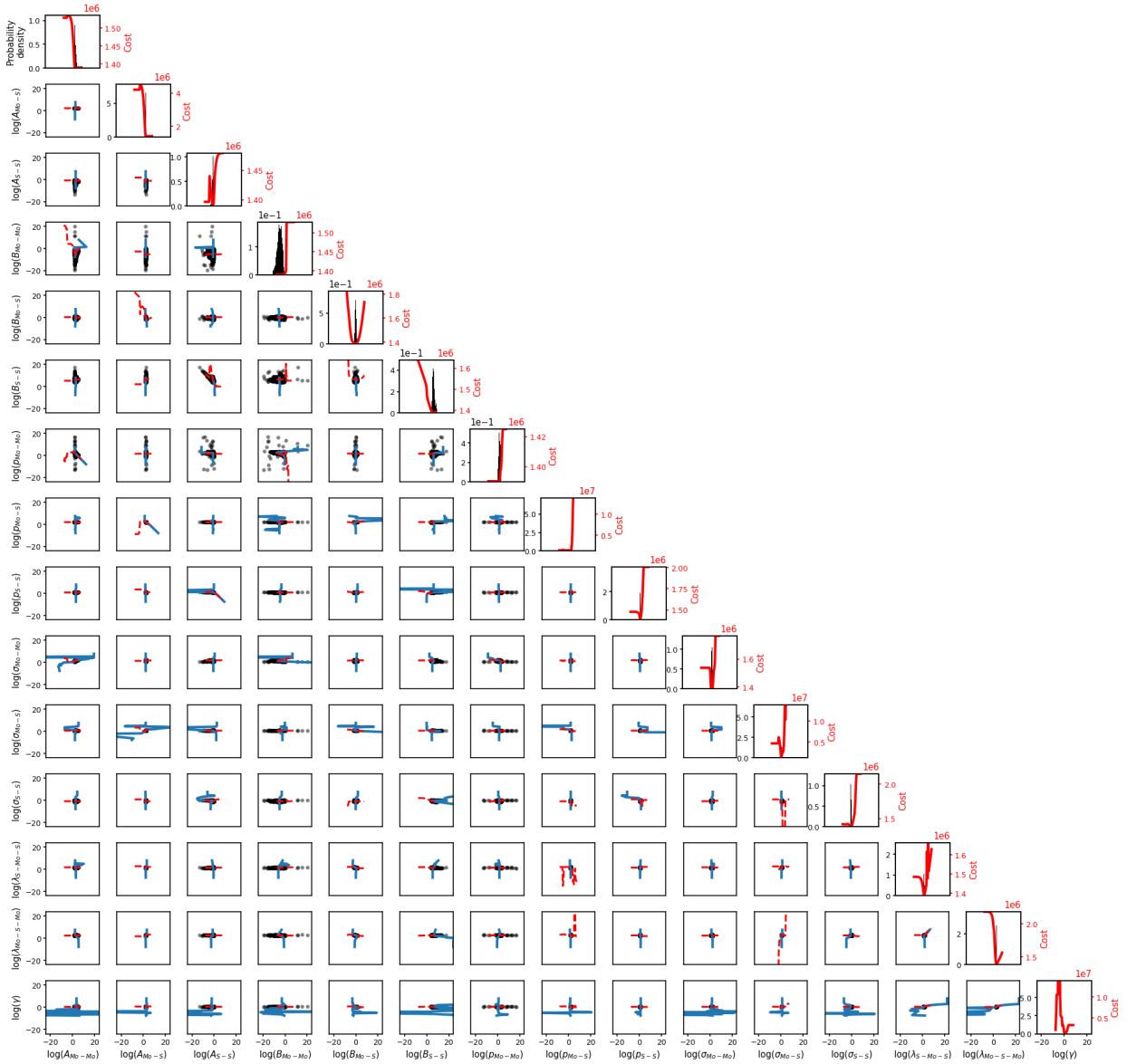


Figure A.8 Sampling Temperature $1.71 \times 10^{-3}\ T_0$

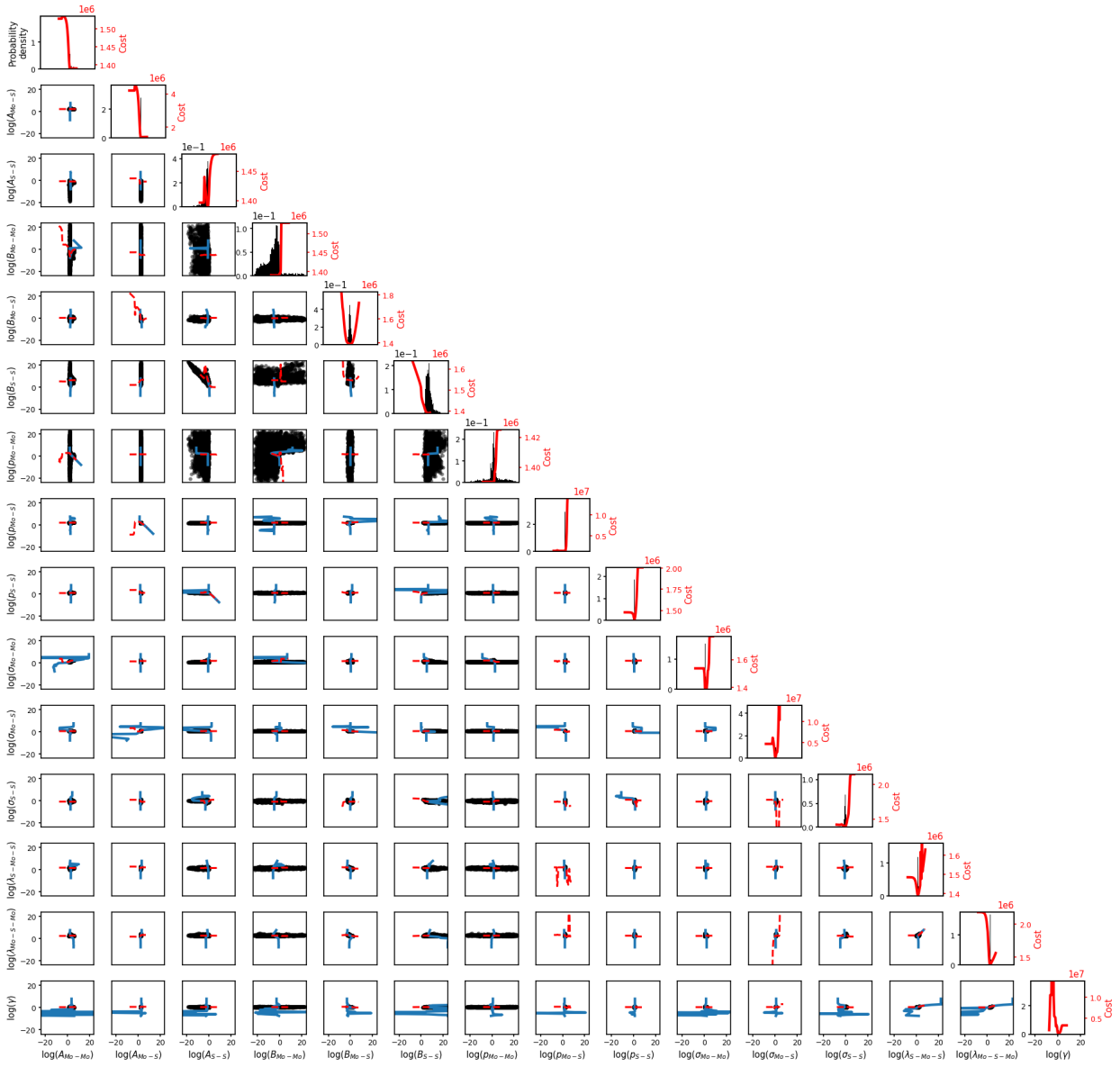


Figure A.9 Sampling Temperature 5.40×10^{-2} T_0

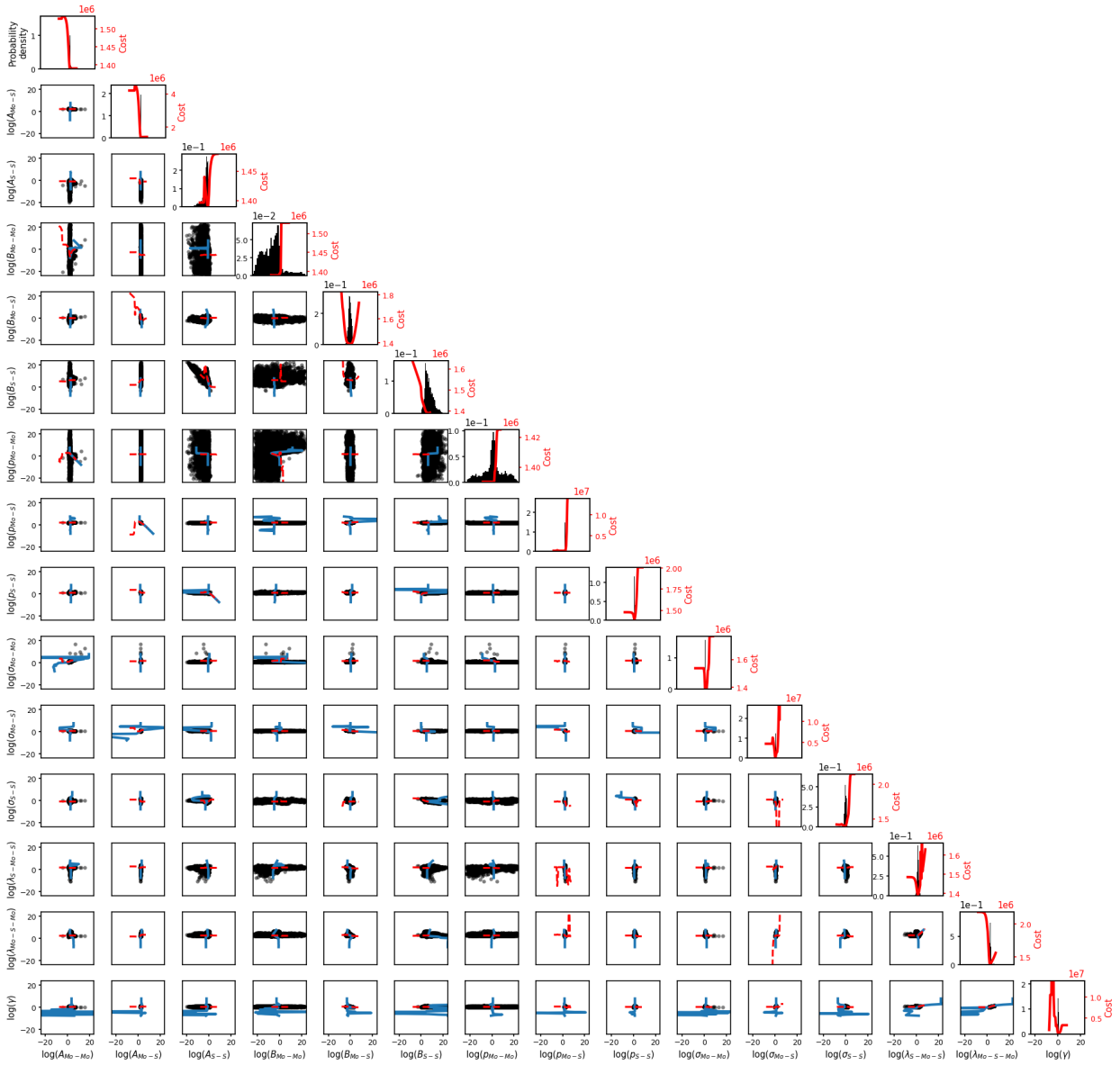


Figure A.10 Sampling Temperature $1.71 \times 10^{-2}\ T_0$

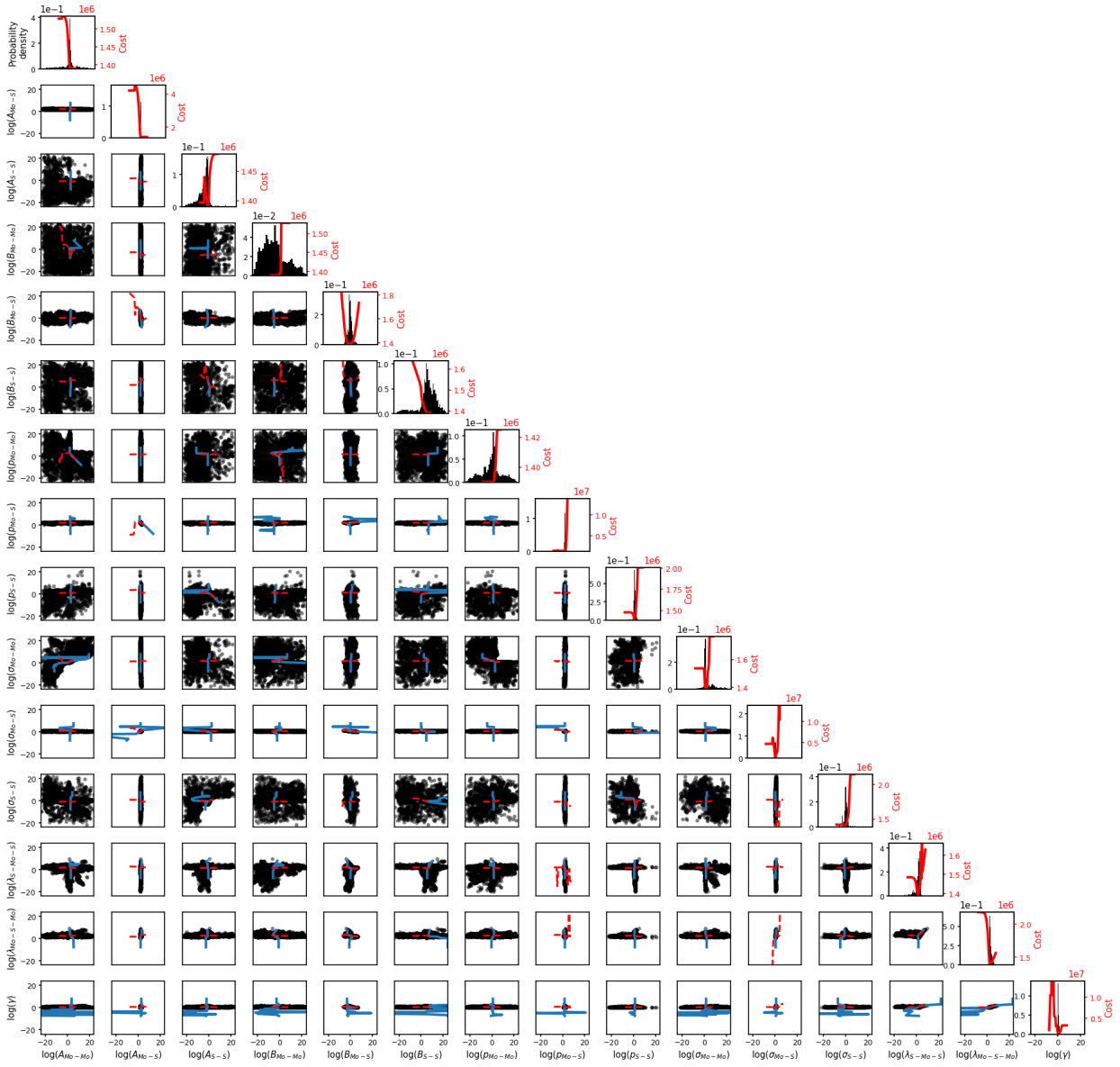


Figure A.11 Sampling Temperature 5.40×10^{-1} T_0

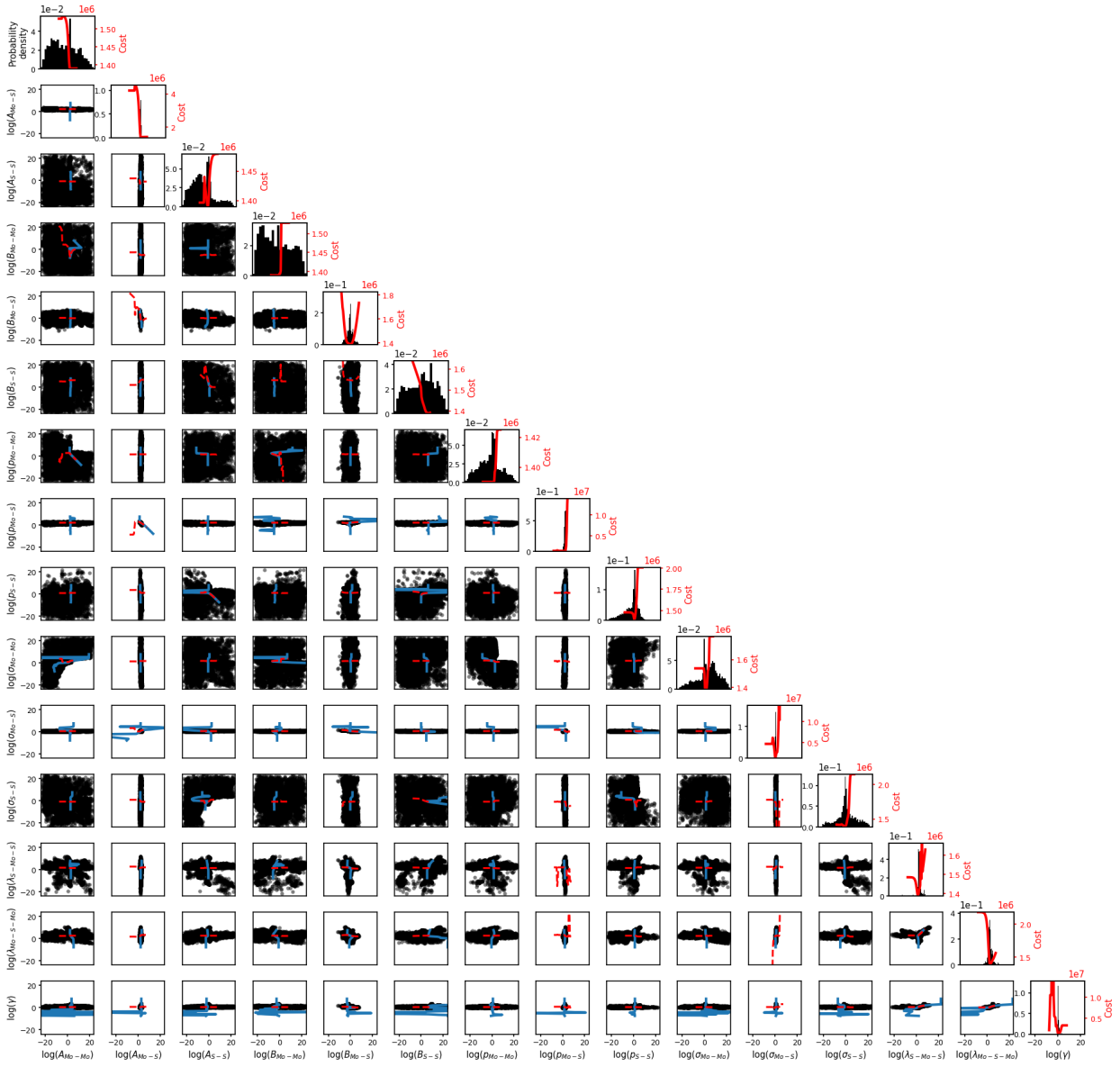


Figure A.12 Sampling Temperature 1.71×10^{-1} T_0

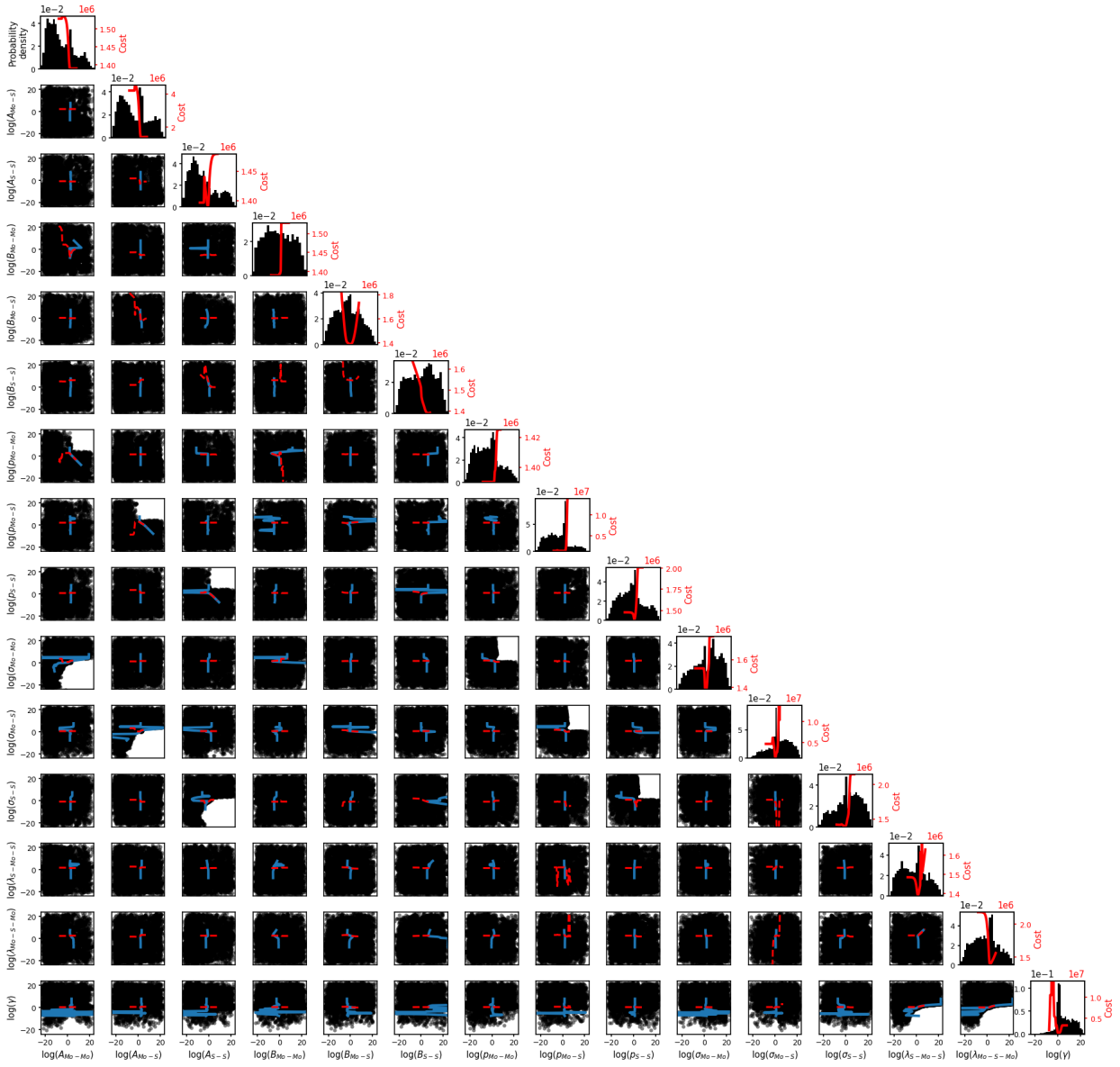


Figure A.13 Sampling Temperature 5.40 T_0

Bibliography

[1] E. B. Tadmor and R. E. Miller, *Modeling Materials: Continuum, Atomistic and Multiscale Techniques* (Cambridge University Press, 2011).

[2] M. Wen, S. N. Shirodkar, P. Plecháč, E. Kaxiras, R. S. Elliott, and E. B. Tadmor, "A force-matching Stillinger-Weber potential for MoS_2: Parameterization and Fisher information theory based sensitivity analysis," Journal of Applied Physics **122,** 244301 (2017).

[3] S. Longbottom and P. Brommer, "Uncertainty quantification for classical effective potentials: an extension to potfit," Modelling and Simulation in Materials Science and Engineering **27,** 044001 (2019).

[4] S. L. Frederiksen, K. W. Jacobsen, K. S. Brown, and J. P. Sethna, "Bayesian Ensemble Approach to Error Estimation of Interatomic Potentials," Physical Review Letters **93,** 165501 (2004).

[5] M. Wen and E. B. Tadmor, "Uncertainty quantification in molecular simulations with dropout neural network potentials," npj Computational Materials 6 (2020).

[6] A. Chernatynskiy, S. R. Phillpot, and R. LeSar, "Uncertainty Quantification in Multiscale Simulation of Materials: A Prospective," Annual Review of Materials Research **43,** 157–182 (2013).

[7] K. S. Brown and J. P. Sethna, "Statistical mechanical approaches to models with many poorly known parameters," Physical Review E **68,** 021904 (2003).

[8] M. K. Transtrum, B. B. Machta, and J. P. Sethna, "Why are nonlinear fits so challenging?," Physical Review Letters **104,** 060201 (2010), arXiv: 0909.3884.

[9] M. K. Transtrum, B. B. Machta, K. S. Brown, B. C. Daniels, C. R. Myers, and J. P. Sethna, "Perspective: Sloppiness and emergent theories in physics, biology, and beyond," The Journal of Chemical Physics **143,** 010901 (2015).

[10] D. W. Brenner, "The Art and Science of an Analytic Potential," physica status solidi (b) **217,** 23–40 (2000).

[11] E. B. Tadmor, R. S. Elliott, J. P. Sethna, R. E. Miller, and C. A. Becker, "The potential of atomistic simulations and the Knowledgebase of Interatomic Models," JOM **63,** 17–17 (2011).

[12] F. Ercolessi and J. B. Adams, "Interatomic Potentials from First-Principles Calculations: The Force-Matching Method," EPL (Europhysics Letters) **26,** 583 (1994).

[13] J. M. Hughes, M. F. Horstemeyer, R. Carino, N. Sukhija, W. B. Lawrimore, S. Kim, and M. I. Baskes, "Hierarchical Bridging Between Ab Initio and Atomistic Level Computations: Sensitivity and Uncertainty Analysis for the Modified Embedded-Atom Method (MEAM) Potential (Part B)," JOM **67,** 148–153 (2015).

[14] P. Angelikopoulos, C. Papadimitriou, and P. Koumoutsakos, "Bayesian uncertainty quantification and propagation in molecular dynamics simulations: A high performance computing framework," The Journal of Chemical Physics **137,** 144103 (2012).

[15] F. Rizzi, R. E. Jones, B. J. Debusschere, and O. M. Knio, "Uncertainty quantification in MD simulations of concentration driven ionic flow through a silica nanopore. II. Uncertain potential parameters," The Journal of Chemical Physics **138,** 194105 (2013).

[16] R. A. Messerly, M. R. Shirts, and A. F. Kazakov, "Uncertainty quantification confirms unreliable extrapolation toward high pressures for united-atom Mie λ-6 force field," The Journal of Chemical Physics **149,** 114109 (2018).

[17] M. Vohra, A. Y. Nobakht, S. Shin, and S. Mahadevan, "Uncertainty quantification in non-equilibrium molecular dynamics simulations of thermal transport," International Journal of Heat and Mass Transfer **127,** 297–307 (2018).

[18] M. Vohra and S. Mahadevan, "Discovering the active subspace for efficient UQ of molecular dynamics simulations of phonon transport in silicon," International Journal of Heat and Mass Transfer **132,** 577–586 (2019).

[19] G. Dhaliwal, P. B. Nair, and C. V. Singh, "Uncertainty analysis and estimation of robust AIREBO parameters for graphene," Carbon **142,** 300–310 (2019).

[20] G. Dhaliwal, P. B. Nair, and C. V. Singh, "Uncertainty and sensitivity analysis of mechanical and thermal properties computed through Embedded Atom Method potential," Computational Materials Science **166,** 30–41 (2019).

[21] L. De Simon, M. Iglesias, B. Jones, and C. Wood, "Quantifying uncertainty in thermophysical properties of walls by means of Bayesian inversion," Energy and Buildings **177,** 220–245 (2018).

[22] F. Rizzi, H. N. Najm, B. J. Debusschere, K. Sargsyan, M. Salloum, H. Adalsteinsson, and O. M. Knio, "Uncertainty Quantification in MD Simulations. Part II: Bayesian Inference of Force-Field Parameters," Multiscale Modeling & Simulation **10,** 1460–1492 (2012).

[23] M. Cools-Ceuppens and T. Verstraelen, Ph.D. thesis, Ghent University, 2017.

[24] R. A. Messerly, T. A. Knotts, and W. V. Wilding, "Uncertainty quantification and propagation of errors of the Lennard-Jones 12-6 parameters for n-alkanes," The Journal of Chemical Physics **146,** 194110 (2017).

[25] M. A. Tschopp, B. Chris Rinderspacher, S. Nouranian, M. I. Baskes, S. R. Gwaltney, and M. F. Horstemeyer, "Quantifying Parameter Sensitivity and Uncertainty for Interatomic Potential Design: Application to Saturated Hydrocarbons," ASCE-ASME J Risk and Uncert in Engrg Sys Part B Mech Engrg 4 (2018).

[26] A. Mishra, S. Hong, P. Rajak, C. Sheng, K.-i. Nomura, R. K. Kalia, A. Nakano, and P. Vashishta, "Multiobjective genetic training and uncertainty quantification of reactive force fields," npj Computational Materials **4,** 1–7 (2018).

[27] F. Lu, H. Wang, D. Yan, D. Zhang, and W. Xiao, "Application of profile likelihood function to the uncertainty analysis of hydrometeorological extreme inference," Science China Technological Sciences **56,** 3151–3160 (2013).

[28] S. Chen, Y. Li, J. Shin, and T. Kim, "Constructing confidence intervals of extreme rainfall quantiles using Bayesian, bootstrap, and profile likelihood approaches," Science China Technological Sciences **59,** 573–585 (2016).

[29] Y. Wu, L. Xue, Y. Liu, and L. Ren, "Uncertainty assessment of extreme flood estimation in the Dongting Lake basin, China," Hydrology Research **50,** 1162–1176 (2019).

[30] A. F. Brouwer and M. C. Eisenberg, "The underlying connections between identifiability, active subspaces, and parameter space dimension reduction," arXiv preprint arXiv:1802.05641 (2018).

[31] A. White, M. Tolman, H. D. Thames, H. R. Withers, K. A. Mason, and M. K. Transtrum, "The Limitations of Model-Based Experimental Design and Parameter Estimation in Sloppy Systems," PLOS Computational Biology **12,** e1005227 (2016).

[32] R. N. Gutenkunst, J. J. Waterfall, F. P. Casey, K. S. Brown, C. R. Myers, and J. P. Sethna, "Universally Sloppy Parameter Sensitivities in Systems Biology Models," PLOS Computational Biology **3,** 1–8 (2007).

[33] J. J. Waterfall, F. P. Casey, R. N. Gutenkunst, K. S. Brown, C. R. Myers, P. W. Brouwer, V. Elser, and J. P. Sethna, "Sloppy-Model Universality Class and the Vandermonde Matrix," Phys. Rev. Lett. **97,** 150601 (2006).

[34] M. K. Transtrum and P. Qiu, "Model Reduction by Manifold Boundaries," Physical Review Letters **113,** 098701 (2014).

[35] B. B. Machta, R. Chachra, M. K. Transtrum, and J. P. Sethna, "Parameter Space Compression Underlies Emergent Theories and Predictive Models," Science **342,** 604–607 (2013).

[36] J. E. Jeong, Q. Zhuang, M. K. Transtrum, E. Zhou, and P. Qiu, "Experimental design and model reduction in systems biology," Quantitative Biology **6,** 287–306 (2018).

[37] M. K. Transtrum and P. Qiu, "Bridging Mechanistic and Phenomenological Models of Complex Biological Systems," PLOS Computational Biology **12,** e1004915 (2016).

[38] B. K. Mannakee, A. P. Ragsdale, M. K. Transtrum, and R. N. Gutenkunst, in *Uncertainty in Biology*, Vol. 17 of *Studies in Mechanobiology, Tissue Engineering and Biomaterials* (Springer International Publishing, 2016), p. 271–299.

[39] M. K. Transtrum and P. Qiu, "Optimal experiment selection for parameter estimation in biological differential equation models," BMC Bioinformatics **13,** 181 (2012).

[40] M. K. Transtrum, A. T. Sarić, and A. M. Stanković, "Information geometry approach to verification of dynamic models in power systems," IEEE Transactions on Power Systems **33,** 440–450 (2017).

[41] R. Gutenkunst, Ph.D. thesis, Cornell University, 2007, accepted: 2007-08-29T17:25:45Z.

[42] M. Wen, J. Li, P. Brommer, R. S. Elliott, J. P. Sethna, and E. B. Tadmor, "A KIM-compliantpotfitfor fitting sloppy interatomic potentials: application to the EDIP model for silicon," Modelling and Simulation in Materials Science and Engineering **25,** 014001 (2016).

[43] Y. Zuo *et al.*, "Performance and cost assessment of machine learning interatomic potentials," The Journal of Physical Chemistry A **124,** 731–745 (2020).

[44] V. L. Deringer, M. A. Caro, and G. Csányi, "Machine learning interatomic potentials as emerging tools for materials science," Advanced Materials **31,** 1902765 (2019).

[45] M. Wen and E. B. Tadmor, "Hybrid neural network potential for multilayer graphene," Physical Review B **100,** 195419 (2019).

[46] A. van den Bos, in *Parameter Estimation for Scientists and Engineers* (John Wiley & Sons, Ltd, 2007), p. 45–97.

[47] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes*, 3rd ed. ed. (Cambridge University Press, Cambridge, UK; New York, 2007).

[48] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov chain Monte Carlo in practice*, 1st ed. ed. (Chapman & Hall, London; New York, 1996).

[49] C. H. LaMont and P. A. Wiggins, "Correspondence between thermodynamics and inference," Physical Review E **99,** 052140 (2019).

[50] E. T. Jaynes, "Information theory and statistical mechanics," Physical review **106,** 620 (1957).

[51] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, "emcee: The MCMC Hammer," Publications of the Astronomical Society of the Pacific **125,** 306–312 (2013), arXiv: 1202.3665.

[52] J. Goodman and J. Weare, "Ensemble samplers with affine invariance," Communications in Applied Mathematics and Computational Science **5,** 65–80 (2010).

[53] W. Vousden, W. M. Farr, and I. Mandel, "Dynamic temperature selection for parallel-tempering in Markov chain Monte Carlo simulations," Monthly Notices of the Royal Astronomical Society **455,** 1919–1937 (2016), arXiv: 1501.05823.

[54] A. Gelman and D. B. Rubin, "Inference from Iterative Simulation Using Multiple Sequences," Statistical Science **7,** 457–472 (1992), zbl: 06853057.

[55] S. P. Brooks and A. Gelman, "General Methods for Monitoring Convergence of Iterative Simulations," Journal of Computational and Graphical Statistics **7,** 434–455 (1998).

[56] D. Vats and C. Knudson, "Revisiting the Gelman-Rubin Diagnostic," arXiv:1812.09384 [stat] (2020), arXiv: 1812.09384.

[57] M. J. Bayarri and J. O. Berger, "The Interplay of Bayesian and Frequentist Analysis," Statistical Science **19,** 58–80 (2004).

[58] E.-J. Wagenmakers, M. Lee, T. Lodewyckx, and G. J. Iverson, in *Bayesian Evaluation of Informative Hypotheses*, *Statistics for Social and Behavioral Sciences* (Springer, 2008), p. 181–207.

[59] K. B. Athreya and S. N. Lahiri, *Measure theory and probability theory* (Springer Science & Business Media, 2006).

[60] S. R. Cole, H. Chu, and S. Greenland, "Maximum Likelihood, Profile Likelihood, and Penalized Likelihood: A Primer," American Journal of Epidemiology **179,** 252–260 (2014).

[61] W. A. Rolke, A. M. López, and J. Conrad, "Limits and confidence intervals in the presence of nuisance parameters," Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **551,** 493–503 (2005).

[62] Y. Kurniawan, "profile_likelihood,", https://gitlab.com/yonatank93/profile_likelihood, 2021.

[63] M. K. Transtrum and J. P. Sethna, "Improvements to the Levenberg-Marquardt algorithm for nonlinear least-squares minimization," arXiv:1201.5885 [physics] (2012), arXiv: 1201.5885.

[64] M. K. Transtrum, "MBAM GitHub repository,", https://github.com/mktranstrum/MBAM, 2018.

[65] R. S. Elliott and E. B. Tadmor, "Knowledgebase of Interatomic Models (KIM) Application Programming Interface (API),", https://openkim.org/kim-api, 2011.

[66] R. S. Elliott, "Efficient 'universal' shifted Lennard-Jones model for all KIM API supported species developed by Elliott and Akerson (2015) v003,", OpenKIM, https://doi.org/10.25950/962b4967, 2018.

[67] R. S. Elliott, "Efficient multi-species Lennard-Jones model with truncated or shifted cutoff v003,", OpenKIM, https://doi.org/10.25950/962b4967, 2018.

[68] J. E. Jones, "On the Determination of Molecular Fields. I. From the Variation of the Viscosity of a Gas with Temperature," Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences **106,** 441–462 (1924).

[69] J. E. Jones, "On the Determination of Molecular Fields. II. From the Equation of State of a Gas," Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences **106,** 463–477 (1924).

[70] J. E. Lennard-Jones, "On the Forces between Atoms and Ions," Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences **109,** 584–497 (1925).

[71] R. S. Elliott, "Morse potential (shifted) for Ni by Girifalco and Weizer (1959) using a high-accuracy cutoff distance v004,", OpenKIM, https://doi.org/10.25950/fc55a3bb, 2020.

[72] R. S. Elliott and Y. Afshar, "Morse pair potential shifted to zero energy at cutoff separation v004,", OpenKIM, https://doi.org/10.25950/fc55a3bb, 2020.

[73] L. A. Girifalco and V. G. Weizer, "Application of the Morse Potential Function to Cubic Metals," Physical Review **114,** 687–690 (1959).

[74] D. S. Karls, "Potential energy and atomic forces of periodic, non-orthogonal cell of Si atoms in a perturbed bcc structure v003,", OpenKIM, https://openkim.org/cite/TE_006970922000_003, 2019.

[75] D. S. Karls, "Potential energy and atomic forces of periodic, non-orthogonal cell of atoms v003,", OpenKIM, https://doi.org/10.25950/c3dca28e, 2019.

[76] D. S. Karls, "Potential energy and atomic forces of periodic, non-orthogonal cell of Si atoms in a perturbed random structure v003,", OpenKIM, https://openkim.org/cite/TE_047848074253_003, 2019.

[77] M. Wen, "Modified Stillinger-Weber potential (MX2) for monolayer MoS2 developed by Wen et al. (2017) v001,", OpenKIM, https://doi.org/10.25950/eeedbbc4, 2018.

[78] M. Wen, "Stillinger-Weber Model Driver for Monolayer MX2 systems v001,", OpenKIM, https://doi.org/10.25950/eeedbbc4, 2018.

[79] A. Chawla, "Three-body cluster potential for Si by Khor and Das Sarma (1988) v000,", OpenKIM, https://doi.org/10.25950/8e5d84c2, 2019.

[80] K. E. Khor and S. Das Sarma, "Proposed universal interatomic potential for elemental tetrahedrally bonded semiconductors," Phys. Rev. B **38,** 3318–3322 (1988).

[81] D. S. Karls, "Environment-Dependent Interatomic Potential (EDIP) model driver v002,", OpenKIM, https://doi.org/10.25950/545ca247, 2018.

[82] D. S. Karls, "EDIP model for Si developed by Justo et al. (1998) v002,", OpenKIM, https://doi.org/10.25950/545ca247, 2018.

[83] M. Z. Bazant and E. Kaxiras, "Modeling of Covalent Bonding in Solids by Inversion of Cohesive Energy Curves," Physical Review Letters **77,** 4370–4373 (1996).

[84] J. a. F. Justo, M. Z. Bazant, E. Kaxiras, V. V. Bulatov, and S. Yip, "Interatomic potential for silicon defects and disordered phases," Physical Review B **58,** 2539–2550 (1998).

[85] M. Wen, "Stillinger-Weber (SW) Model Driver v004,", OpenKIM, https://doi.org/10.25950/e9adf627, 2018.

[86] A. K. Singh, "Stillinger-Weber potential for Si developed by Balamane, Halicioglu and Tiller (1992) v004,", OpenKIM, https://doi.org/10.25950/e9adf627, 2018.

[87] F. H. Stillinger and T. A. Weber, "Computer simulation of local order in condensed phases of silicon," Physical Review B **31,** 5262–5271 (1985).

[88] H. Balamane, T. Halicioglu, and W. A. Tiller, "Comparative study of silicon empirical interatomic potentials," Physical Review B **46,** 2250–2279 (1992).

[89] H. Balamane, T. Halicioglu, and W. A. Tiller, "Vacancy- and adatom-induced $\sqrt{3} \times \sqrt{3}$ reconstructions of the Si(111) surface," Physical Review B **40,** 9999–10001 (1989).

[90] I. J. Perez-arriaga, G. C. Verghese, and F. C. Schweppe, "Selective Modal Analysis with Applications to Electric Power Systems, PART I: Heuristic Introduction," IEEE Transactions on Power Apparatus and Systems **PAS-101,** 3117–3125 (1982).

[91] F. Garofalo, L. Iannelli, and F. Vasca, "PARTICIPATION FACTORS AND THEIR CONNECTIONS TO RESIDUES AND RELATIVE GAIN ARRAY," IFAC Proceedings Volumes **35,** 125–130 (2002).

[92] C. Robert, "On the Jeffreys-Lindley's paradox," arXiv:1303.5973 [stat] (2013), arXiv: 1303.5973.

[93] M. K. Transtrum, B. B. Machta, and J. P. Sethna, "Geometry of nonlinear least squares with applications to sloppy models and optimization," Physical Review E **83,** 036701 (2011).

[94] D. M. Bates and D. G. Watts, "Relative Curvature Measures of Nonlinearity," Journal of the Royal Statistical Society: Series B (Methodological) **42,** 1–16 (1980).