THE EUROPEAN
PHYSICAL JOURNAL C

Efficient sampling of constrained high-dimensional theoretical spaces with machine learning

Jacob Hollingsworth[1,a], **Michael Ratz**[1], **Philip Tanedo**[2], **Daniel Whiteson**[1]

[1] Department of Physics and Astronomy, University of California, Irvine, CA 92697, USA
[2] Department of Physics and Astronomy, University of California, Riverside, CA 92521, USA

Abstract Models of physics beyond the Standard Model often contain a large number of parameters. These form a high-dimensional space that is computationally intractable to fully explore. Experimental results project onto a subspace of parameters that are consistent with those observations, but mapping these constraints to the underlying parameters is also typically intractable. Instead, physicists often resort to scanning small subsets of the full parameter space and testing for experimental consistency. We propose an alternative approach that uses generative models to significantly improve the computational efficiency of sampling high-dimensional parameter spaces. To demonstrate this, we sample the constrained and phenomenological Minimal Supersymmetric Standard Models subject to the requirement that the sampled points are consistent with the measured Higgs boson mass. Our method achieves orders of magnitude improvements in sampling efficiency compared to a brute force search.

1 Introduction

Models of physics beyond the Standard Model often feature many new parameters that are unknown a priori and may only be determined by experiment. However, experimental constraints are not trivial to apply, as they often are expressed in terms of weak scale observables rather than the theory's fundamental parameters. While it is often straightforward (if computationally expensive) to calculate the weak scale observables from the parameters, the inverse problem is typically intractable. That is, weak scale constraints do not allow for a trivial reduction of the dimensionality of the theory space.

The standard approach is to numerically scan over the theoretical parameters and reject those that are not consis-

tent with experimental data. However, the number of samples required for a brute-force search of the parameter space increases exponentially with its dimension. Thus, particle physicists studying models of new physics are often faced with a computationally intractable task. One may pragmatically restrict to a more tractable subset of parameters based on theoretical prejudice. The danger of this approach is that one may miss viable parameters that are both consistent with experimental observations and generate novel phenomenology.

The Minimal Supersymmetric Standard Model (MSSM) is a well-known example of a new physics model with a large number of free parameters. Most of these parameters are the masses and couplings of the supersymmetric partners of Standard Model particles [1]. This overwhelming dimensionality prohibits a fully general survey of the parameter space. Studies of the MSSM typically restrict to theoretically motivated subspaces [2–13]. These include the 4+1 dimensional constrained MSSM (cMSSM) as well as the 19 dimensional phenomenological MSSM (pMSSM) [14,15]. However, even these reduced spaces are difficult to scan using a brute-force search.

High dimensionality is not the only challenge when scanning the parameters of the MSSM. The fundamental parameters of the theory are defined at some high energy scale and must be evolved to the energy scale of the experiment. This evolution requires one to solve the coupled (RGEs) for the high-scale parameters over many orders of magnitude to the weak scale. The computational cost of RGE running and calculating experimental observables for a single set of parameters is expensive.

Many recent scans have incorporated machine learning in some capacity to decrease the computational burden of brute-force searching these spaces [7,12,13]. These use various machine learning models to learn the *forward* problem of determining weak scale properties given high-scale parameters. This bypasses the need to perform RGE run-

ning and weak scale computations, however one is still faced with the challenge of doing a brute-force search over a high-dimensional parameter space. Machine learning models for the forward problem are thus only a constant improvement in computational time compared to the exponential dependence on the dimension of the space.

In this work, we introduce two methods to efficiently sample high-dimensional parameter spaces subject to constraints at the weak scale. We test these frameworks by sampling regions of the cMSSM and pMSSM parameter spaces that admit a Higgs mass consistent with its experimental value [16,17]. The first uses a deep neural network to machine-learn the likelihood of an event satisfying this constraint and then samples this likelihood using Hamiltonian Monte Carlo (HMC). The second trains a generative model known as a normalizing flow. We analyze the performance of these frameworks by determining the fraction of generated samples that survive the chosen constraint and compare to the performance of random sampling.

These methods allow us to directly and quickly generate points in the parameter space that admit a consistent Higgs mass. By solving the inverse problem of sampling high-scale parameters given weak scale properties, we aim to minimize inefficiencies that arise in a brute-force search.

Our presentation is a proof of concept for these generative models and is encouraging for practical applications. For example, the ability to efficiently scan the MSSM parameter space makes it much easier to determine the high-scale parameters that are consistent with a new particle's mass and width if a sparticle is discovered. Alternatively, a trained generative model may permit scans over parameters that are consistent with experimental observations to search for specific theoretical features that one may wish to study, for example: gauge coupling unification, a particular type of dark matter particle, or low fine-tuning measures.

As a demonstration of the efficiency of the generative models, we scan the cMSSM and pMSSM parameter spaces for points that produce the Higgs mass and that saturate the observed dark matter relic density, requiring [18,19]

$$122 \text{ GeV} < m_h < 128 \text{ GeV},$$
$$0.08 < \Omega_{\text{DM}} h^2 < 0.14.$$

In this study, the generative models have been trained for consistency with the Higgs mass, not the relic density. We compare a brute-force scan using random sampling to a generative model that has been trained to sample points that admit a consistent Higgs mass. We show that the generative models dramatically increase the sampling efficiency of this scan.

Table 1 Parameter bounds in the cMSSM scan, following Ref. [2]. A uniform prior is used for all parameters except A_0, where we uniformly sample A_0/m_0

Parameter	Domain	Description
m_0	[0, 10] TeV	Universal scalar mass
$m_{1/2}$	[0, 10] TeV	Universal gaugino mass
A_0	$[-6m_0, 6m_0]$ TeV	Universal trilinear coupling
$\tan \beta$	[1.5, 50]	Ratio of Higgs VEVs

2 Methods

2.1 Data generation

The cMSSM contains 4 continuous parameters defined at the Grand Unified Theory (GUT) scale and 1 discrete sign parameter. These are the universal scalar mass m_0, the universal gaugino mass $M_{1/2}$, universal trilinear coupling A_0, the ratio of Higgs vacuum expectation values $\tan \beta$, and the sign of μ. The pMSSM is the most general subspace of the MSSM that admits first and second generation universality, no new sources of CP violation, and no flavor changing neutral currents [15]. Parameters of the pMSSM are defined at the electroweak (EW) scale. The full list parameters of the pMSSM are listed as part of Table 2.

Our datasets are formed by uniform random sampling within bounded regions of the parameter space: cMSSM parameters are sampled at the GUT scale and pMSSM parameters are sampled at the EW scale. Bounds are listed for the cMSSM and the pMSSM in Tables 1 and 2, respectively [2,9], and are chosen to cover large volumes of the parameter space that are sensitive to modern collider experiments. For the cMSSM, we fix sign(μ) = 1. We sample approximately 1.5×10^6 datapoints in the cMSSM and approximately 1.95×10^7 datapoints in the pMSSM. Once sampled, we calculate Higgs masses and relic densities with micrOMEGAs, which internally uses the spectrum generator SoftSUSYv4.1.0 [20,21].

We apply two theoretical constraints: (i) consistent electroweak symmetry breaking and (ii) the positivity of all squared masses. In addition to these, we also require that SoftSUSY converges. We do not require that the lightest supersymmetric particle is neutral.

The theoretical uncertainty in the Higgs mass is significantly larger than its experimental uncertainty [22]. We take the uncertainty in the Higgs mass calculations to be $\sigma_{m_h} = 3$ GeV for all points in the data set [2,9].

Table 2 Parameter bounds in the pMSSM scan, following Ref. [9]. A uniform prior is used for all parameters. "Left-handed" and "right-handed" are abbreviated by l.h. and r.h., respectively

Parameter	Domain	Description
$|M_1|$	[.05, 4] TeV	Bino mass
$|M_2|$	[.1, 4] TeV	Wino mass
M_3	[.4, 4] TeV	Gluino mass
$|\mu|$	[.1, 4] TeV	Bilinear Higgs mass
$|A_t|$	[0, 4] TeV	Trilinear top coupling
$|A_b|$	[0, 4] TeV	Trilinear bottom coupling
$|A_\tau|$	[0, 4] TeV	Trilinear τ coupling
M_A	[.1, 4] TeV	Pseudo-scalar Higgs mass
$m_{\tilde{L}_1}$	[.1, 4] TeV	1st gen. l.h. slepton mass
$m_{\tilde{e}_1}$	[.1, 4] TeV	1st gen. r.h. slepton mass
$m_{\tilde{L}_3}$	[.1, 4] TeV	3rd gen. l.h. slepton mass
$m_{\tilde{e}_3}$	[.1, 4] TeV	3rd gen. r.h. slepton mass
$m_{\tilde{Q}_1}$	[.4, 4] TeV	1st gen. l.h. squark mass
$m_{\tilde{u}_1}$	[.4, 4] TeV	1st gen. r.h. u-type squark mass
$m_{\tilde{d}_1}$	[.4, 4] TeV	1st gen. r.h. d-type squark mass
$m_{\tilde{Q}_3}$	[.2, 4] TeV	3rd gen. l.h. squark mass
$m_{\tilde{u}_3}$	[.2, 4] TeV	3rd gen. r.h. u-type squark mass
$m_{\tilde{d}_3}$	[.2, 4] TeV	3rd gen. r.h. d-type squark mass
$\tan\beta$	[1, 60]	Ratio of Higgs VEVs

2.2 Neural network

We train the neural network by assigning all points in the dataset a likelihood

$$L(\theta) = \begin{cases} 1 & |m_h(\theta) - m_{h,\text{exp}}| < \sigma_{m_h}, \\ 0 & \text{otherwise}, \end{cases} \quad (1)$$

where we ignore a normalization constant. All data points that fail the theoretical constraints are assigned a likelihood of zero.

We use a deep neural network to learn the function $L(\theta)$ [23]. This has two benefits. First, it greatly reduces the time required to evaluate the likelihood of a point. Second, it provides a differentiable interpolation of $L(\theta)$. In the next section we show that HMC requires many evaluations of the likelihood and its gradients. It thus utilizes the full potential of these benefits.

We train a deep neural network $\hat{L}(\theta)$ to minimize the usual L2 loss function

$$\mathcal{L} = |\hat{L}(\theta) - L(\theta)|^2. \quad (2)$$

We use a training, validation, and testing split of 0.7, 0.15, and 0.15 respectively for both datasets. Batch norm and dropout layers are used in between each hidden layer of the neural network. Backpropagation is performed using the ADAM optimizer [24].

Some of the pMSSM parameters in Table 2 span a disconnected range of positive and negative values, for example M_1, M_2 and μ. We preprocess these parameters by shifting negative values to create a single continuous domain; for example, for μ we shift the negative values by 200 GeV. This has no physical significance and simply prepares the data for input into the neural network. We then standardize each feature. For the cMSSM dataset, we use the feature A_0/m_0 in place of A_0, as this feature is uniformly distributed.

2.3 Hamiltonian Monte Carlo

The Hamiltonian Monte Carlo method is a Markov chain Monte Carlo technique that uses an analog of energy conservation to effectively sample the target distribution [25,26]. To use the method, we first define an auxiliary momentum variable p, where each component is initially drawn from a normal distribution. Next, we define a potential energy function given by

$$V(\theta) = -\log(\hat{L}(\theta)). \quad (3)$$

The kinetic energy function takes the familiar form $T = p^2/2$ where we set the mass to unity, $m = 1$. We then evolve the system from time $t = 0$ to $t = \tau$ according to the Hamiltonian equations of motion

$$\frac{d\theta_i}{dt} = p_i, \qquad \frac{dp_i}{dt} = \frac{\nabla\hat{L}(\theta)}{\hat{L}(\theta)}. \quad (4)$$

We solve these equations using the leap-frog algorithm so that energy is approximately conserved. We take $\theta(\tau)$ as a proposal to add to the Markov chain. The proposal is accepted with probability

$$P = \min\left(1, \frac{e^{-H(\theta(\tau),p(\tau))}}{e^{-H(\theta(0),p(0))}}\right). \quad (5)$$

Energy conservation implies that a solution to the the equations of motion should always yield probability 1. However, a rejection step is necessary because we solve these equations numerically. If $\theta(\tau)$ is rejected, then $\theta(0)$ is added to the Markov chain instead. In the limit of an infinite number of samples, the Markov chain converges to a sample of the distribution $\hat{L}(\theta)$. We seed the Markov chain with a random positive sample from the dataset used to train the neural network. We bound the parameter space with hard walls of infinite potential energy.

2.4 Normalizing flows

It is difficult to draw samples from a complicated distribution in a high-dimensional parameter space. On the other hand, it is easy to draw samples from an equally high-dimensional Gaussian distribution. Normalizing flows is a technique that learns an invertible map f from a simple distribution p_Z to a challenging distribution p_Y. One then creates a set of samples from the challenging distribution by mapping easy-to-generate samples:

$$p_Y(y) = p_Z(f^{-1}(y)) \left| \det \left(\frac{\partial f}{\partial y} \right) \right|^{-1}. \quad (6)$$

The function f depends on a set of parameters Θ which are learned by maximizing the log likelihood of a training set, $\mathcal{X}$. The loss function for this training is thus

$$\mathcal{L}(\mathcal{X}) = -\sum_{y \in \mathcal{X}} \left(\log \left(p_Z(f^{-1}(y)) \right) - \log \left| \det \left(\frac{\partial f}{\partial y} \right) \right| \right).$$

It is helpful to construct f to be the composition of n successive maps, $f = f_n \circ \cdots \circ f_1$ [23]. Defining $z_{i+1} = f_i(z_i)$ and identifying $y = z_{n+1}$ yields the loss function

$$\mathcal{L}(\mathcal{X}) = -\sum_{y \in \mathcal{X}} \left(\log (p_Z(z_1)) - \sum_{i=1}^{n} \log \left| \det \left(\frac{\partial z_{i+1}}{\partial z_i} \right) \right| \right).$$

We choose the f_i to be autoregressive transformations. This means that the parameters Θ_i^k that define the function f_i acting on the kth feature z_i^k depends only on the first $(k-1)$ features $z_i^1, \ldots, z_i^{k-1}$:

$$z_{i+1}^k = f_i \left(z_i^k; \ \Theta_i^k(z_i^{1:k-1}) \right).$$

This structure ensures that the Jacobian matrix $\partial z_{i+1} / \partial z_i$ is lower triangular so that the determinant is simply the product of diagonal elements and may be computed in linear time.

The function $\Theta_i^k \left(z_i^{1:k-1} \right)$ can be represented efficiently with a Masked Autoencoder for Distribution Estimation (MADE) [27]. MADE networks turn off specific internal weights of the neural network so that the autoregressive property is enforced, allowing one neural network to output all model parameters rather than performing a sequential loop over features.

For our application, we choose f_i to be rational-quadratic neural spline flows with autoregressive layers [28]. These are piece-wise monotonic functions defined as the ratio of two quadratic functions on the interval $[-B, B]$, with $K+1$ knots determining the boundaries between bins. Outside of this interval, the transformation is defined to be the identity. These transformations are parameterized by $3K-1$ parameters for each feature, which are K bin heights, K bin widths, and $K-1$ positive derivative values at the knots, as the derivatives are set to 1 at $-B$ and B to ensure a continuous derivative over the domain. Permutation layers are included between rational-quadratic transformation layers. We implement the normalizing flow using the Python package *nflows* [28].

3 Results

We analyze the performance of these generative frameworks on the cMSSM and pMSSM datasets described above. The cMSSM is low dimensional and can be scanned relatively well with brute-force search. Thus, we view the cMSSM as a test for the generation methods and the pMSSM as a more practical application. We present the results for the neural network with HMC as well as the normalizing flow side by side. For each method, we generate a dataset of 4×10^5 datapoints.

We present histograms of generated variables to confirm that the distribution of theory parameters is not biased by our generative framework. We also present histograms of m_h to ensure that our generative models sample within the band of permitted Higgs masses and $\Omega_{DM} h^2$ to provide evidence that the distribution of weak scale quantities match, as these are sensitive to higher order correlations in high energy scale parameters. Finally, we report sampling efficiencies, which are defined as the fraction of the dataset that satisfy a constraint. The hyperparameters used for the supervised neural network, Hamiltonian Monte Carlo, and normalizing flow are given in the Appendix for both datasets.

3.1 cMSSM

In Fig. 1, we compare histograms of the cMSSM parameters at the GUT scale. For both generative models, we see very good agreement between the distribution of generated samples and the distribution of randomly sampled points after the Higgs mass constraint is applied. Next, we run the parameters to the weak scale in order to perform the combined search for $\Omega_{DM} h^2$ and m_h. In Fig. 2, we show the distribution of Higgs masses for generated points and randomly sampled points with a rejection step applied. We see that the generative models typically sample within the band of permitted Higgs masses.

As an example application, we show histograms of the dark matter relic density for these datasets in Fig. 3. We see that the distribution over dark matter relic densities from the generative models appear to accurately reflect the same distribution in the dataset after the Higgs mass constraint is applied. We emphasize that because the RGEs are coupled, weak-scale quantities are generally sensitive to higher-order correlations of the GUT scale parameters, and so matching

Fig. 1 Histograms of cMSSM parameters that yield the experimental Higgs mass. We observe good agreement between the random sampling, HMC, and the flow model. Black: Data obtained through random sampling with a uniform prior and rejecting points that do not have a consistent Higgs mass. Magenta: data sampled with HMC. Blue: data sampled from the flow model. No rejection step is applied to generated samples
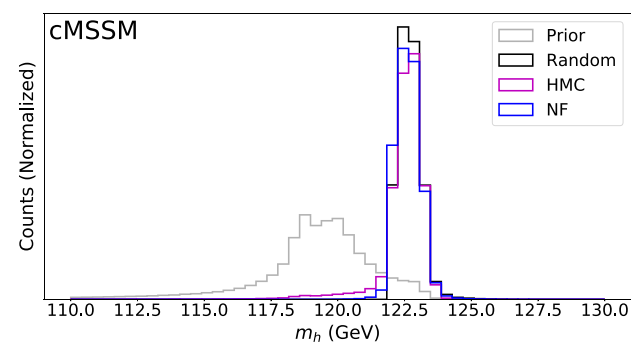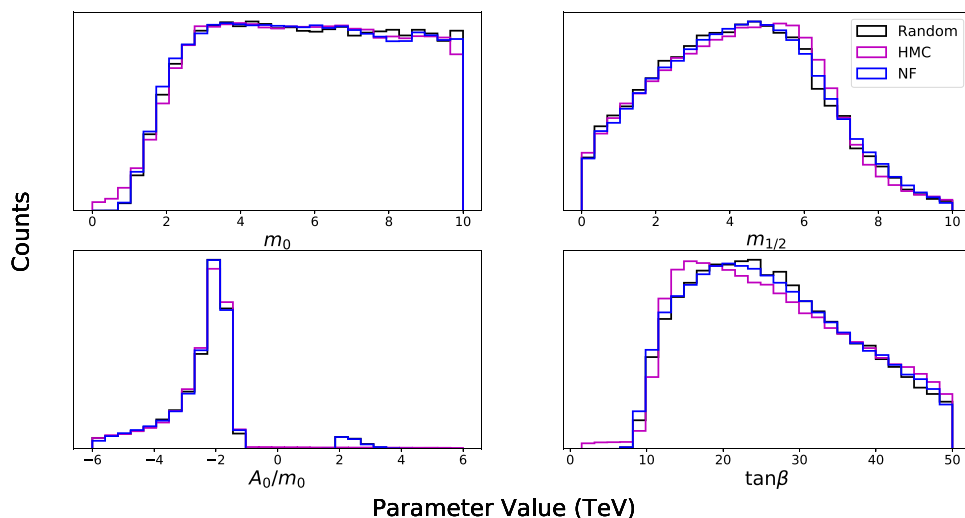

Fig. 2 Histogram of Higgs masses in the cMSSM for different sampling methods. The generative models are seen to mostly sample points consistent with the Higgs mass constraint. Gray: data obtained through random sampling with a uniform prior. Black: the same randomly sampled data, but points that do not have a consistent Higgs mass are rejected. Magenta: data sampled with HMC. Blue: data sampled with the normalizing flow
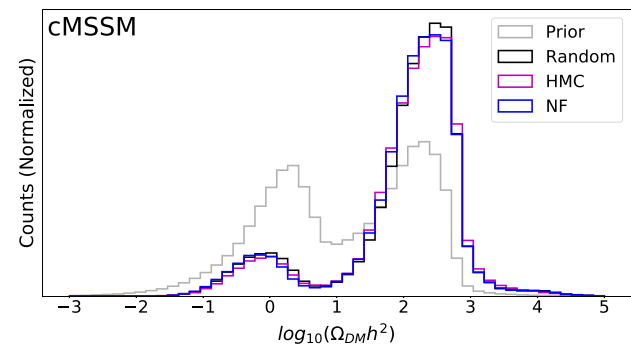
Fig. 3 Histogram of dark matter thermal relic densities in the cMSSM for different sampling methods. We observe that the distributions of the generative models match the distribution of random sampling, providing evidence that the generative models are able to match higher order correlations in GUT scale parameters. Gray: data obtained through random sampling with a uniform prior. Black: the same randomly sampled data, but points that do not have a consistent Higgs mass are rejected. Magenta: data sampled with HMC. Blue: data sampled with the normalizing flow. Generative models have been trained to satisfy the Higgs mass constraint

Table 3 Comparison of sampling efficiency in the cMSSM for several methods and several levels of constraints. We compare a brute force random scan (random), Hamiltonian MC of a neural network trained to learn the m_h constraint (HMC$_{m_h}$), and normalizing flows that incorporate the m_h constraint (NF$_{m_h}$). The constraints applied are theoretical consistency checks (see text), consistency with the experimental Higgs mass and consistency with the Higgs mass and the dark matter relic density ($\Omega_{DM}h^2$)

Constraint	Sampling method		
	Random	HMC$_{m_h}$	NF$_{m_h}$
Theory	0.595	0.859	0.879
Theory $\cap$ m_h	0.0389	0.723	0.796
Theory $\cap$ m_h $\cap$ $\Omega_{DM}h^2$	0.000222	0.00271	0.00456

weak-scale distributions is evidence of matching higher order correlations in the GUT scale parameters. This indicates that the m_h-constrained subspace has been accurately sampled, allowing for an exploration of additional constraints, such as relic density.

In Table 3, we compare various statistical properties of random sampling to those of our generative frameworks trained to satisfy the Higgs mass constraint. The first row shows the sampling efficiency with respect to the theoretical constraints mentioned in Sect. 2.1. We see that samples from the generative models are more likely to pass these constraints, as points with a consistent Higgs mass necessarily satisfy the theoretical constraints. The second row shows the sampling efficiency with respect to the Higgs mass constraint. Predictably, the generative models have significantly higher sampling efficiencies than random sampling. We also see that the flow model slightly outperforms the HMC sampling method.

The third row shows the sampling efficiencies with respect to the combined Higgs mass and relic density constraint, where the generative models are still trained to only satisfy the Higgs mass constraint. This simulates a scenario
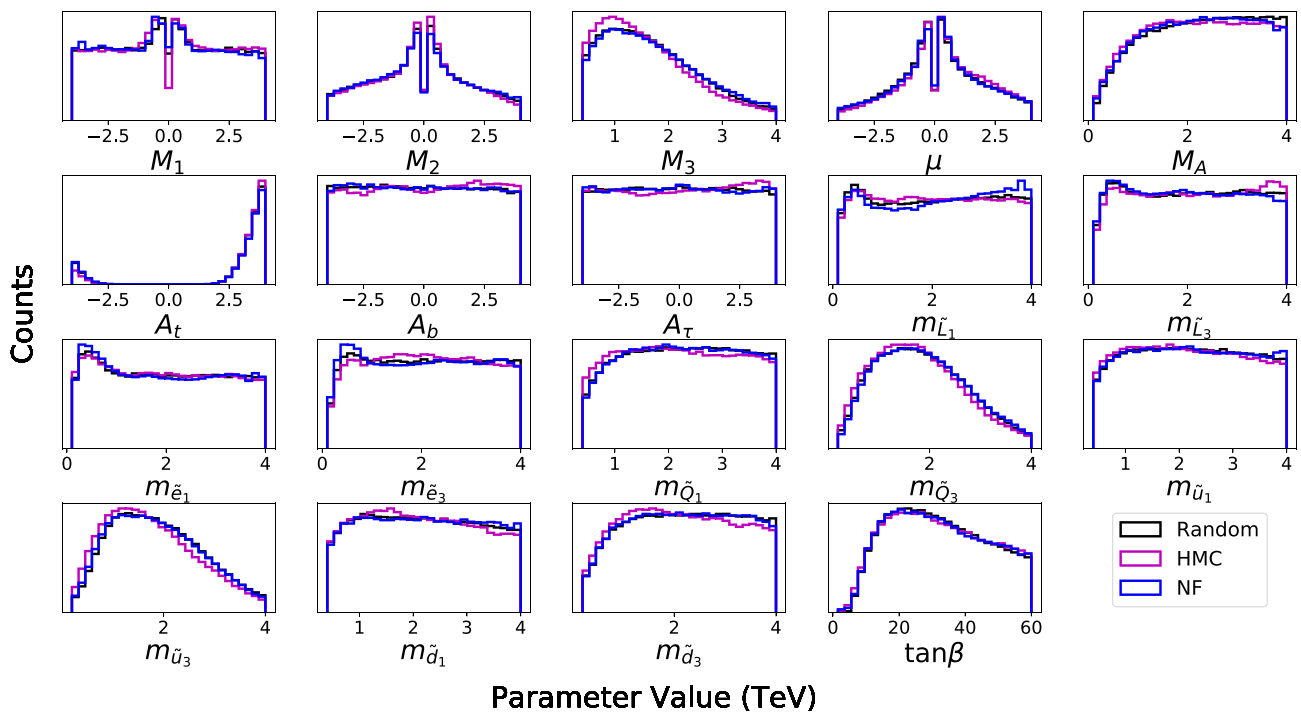
Fig. 4 Histograms of pMSSM parameters that yield the experimental Higgs mass. We observe good agreement between random sampling, HMC, and the flow model. Black: Data obtained through random sam- pling with a uniform prior and rejecting points that do not have a consis- tent Higgs mass. Magenta: data sampled with HMC. Blue: data sampled from the flow model. No rejection step is applied to generated samples

where one would like to study the effect of imposing a new constraint in addition to the constraints that are explicitly trained on. Once again, we see that the generative models have much higher sampling efficiencies, resulting from the high probability that the samples pass the Higgs mass con- straint. We see an increase in sampling efficiency of approx- imately an order of magnitude for both generative frame- works.

3.2 pMSSM

Differences between the generative models appear in the higher-dimensional pMSSM. In Fig. 4, we compare his- tograms of parameters sampled using brute-force search, HMC and the normalizing flow model. Despite the increased dimensionality, we find very good agreement in the distribu- tions of all parameters.

Figures 5 and 6 present histograms of m_h and $\Omega_{DM}h^2$ for the pMSSM. The generative models tend to sample in the band of allowed Higgs masses, with the normalizing flow model matching the brute-force scan well. We see general agreement with the true distribution of dark matter abun- dances for both generative frameworks, though the HMC samples do not match the brute-force distributions as well as those from the flow model.
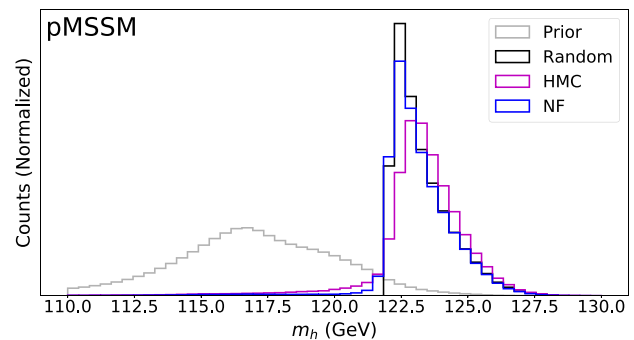

Fig. 5 Histogram of Higgs masses in the pMSSM. The generative models are seen to mostly sample points consistent with the Higgs mass constraint. Gray: data obtained through random sampling with a uni- form prior. Black: the same randomly sampled data, but points that do not have a consistent Higgs mass are rejected. Magenta: data sampled with HMC. Blue: data sampled with the normalizing flow

Table 4 summarizes the performance of our sampling methods in the pMSSM. See Sect. 3.1 for a detailed descrip- tion of the quantities presented in the table. We find that generative models greatly increase the sampling efficiency relative to a brute-force search. In fact, the improvement in sampling efficiency is much greater than that seen in the cMSSM. This is largely due to the poorer performance of a brute-force search in the higher-dimensional pMSSM.
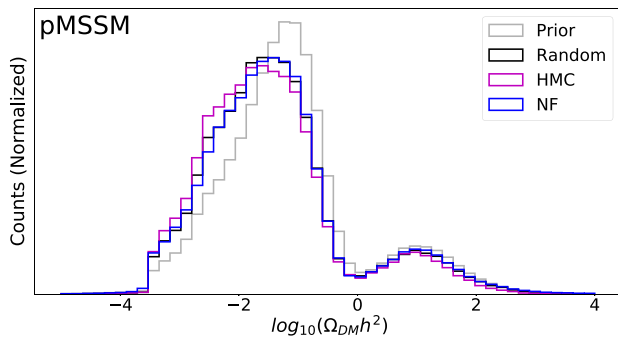
Fig. 6 Histogram of dark matter thermal relic densities in the pMSSM. We observe that the distributions of the generative models match the distribution of random sampling, providing evidence that the generative models are able to match higher order correlations in EW scale parameters. Gray: data obtained through random sampling with a uniform prior. Black: the same randomly sampled data, but points that do not have a consistent Higgs mass are rejected. Magenta: data sampled with HMC. Blue: data sampled with the normalizing flow. Generative models have been trained to satisfy the Higgs mass constraint

Table 4 Comparison of sampling efficiency in the pMSSM for several methods and several levels of constraints. Methods compared are brute force random scan, Hamiltonian MC of a neural network trained to learn the m_h constraint (HMC$_{m_h}$), and normalizing flows that incorporate the m_h constraint (NF$_{m_h}$). Constraints applied are theoretical consistency checks (see text), consistency with the experimental Higgs mass and consistency with the Higgs mass and the dark matter relic density ($\Omega_{DM}h^2$)

Constraint	Sampling method		
	Random	HMC$_{m_h}$	NF$_{m_h}$
Theory	0.479	0.810	0.866
Theory $\cap\, m_h$	0.0189	0.709	0.776
Theory $\cap\, m_h \cap \Omega_{DM}h^2$	0.00165	0.0591	0.0685

4 Conclusion

We implement two generative frameworks that utilize machine learning in order to increase the sampling efficiency of searches in supersymmetric parameter spaces. These sampling methods offer a more efficient way to search the high-dimensional parameter spaces in models of new particle physics. We compare these generative frameworks to the currently used method of a brute-force search, and have seen orders of magnitude of improvement in the sampling efficiency for both parameter spaces considered here. We show that our generative frameworks are able to sample the underlying data distribution without any evidence of bias or mode collapse.

In the cMSSM, both methods significantly outperformed random sampling, with the flow model slightly outperforming HMC. In the pMSSM the flow model significantly outperforms HMC. This is likely due to the larger dimensionality of the pMSSM. In addition to performance benefits, the flow model is also quicker to train and sample, making it clearly favorable to HMC. However, the HMC framework is more complementary to previous works, as it learns the forward problem of determining likelihoods and uses tested Monte Carlo algorithms to sample this likelihood.

Possibilities for future work include incorporating additional constraints into the generative model. In theory, there is no limit to the number of constraints that can be included into either generative model. However, forming an initial dataset for learning may be difficult when the constraints are very strict. A possible remedy is to train generative models with less restrictive constraints which are then used to produce sizable datasets of points that already satisfy many constraints. This new dataset could then be searched to form a training set for a generative model with increasingly restrictive constraints.

Given the ability of the generative machine learning models to efficiently explore high-dimensional parameter spaces, it will be interesting to apply the techniques described here to other problems. For instance, one may identify relations that explain why there is a 'little hierarchy' between the electroweak scale and the scale of soft parameters, which go beyond the focus point scenario [29]. In general, one may be able to identify manifolds of viable points in high-dimensional parameter sets, and explore their geometry.

We have shown promising results in subspaces of the MSSM parameter space. These results apply generally to any high-dimensional parameter space with constraints that are computationally expensive to verify. Another direction for future study may be applications to the parameter spaces of even higher-dimensional models of new physics. This includes potentially relaxing constraints built into the pMSSM parameter space, but could also include applications to non-supersymmetry (SUSY) theories. Finally, one could attempt to further tune the neural network structure and hyperparameters in order to achieve higher sample efficiency than was achieved in this work.

Appendix A

We present the hyperparameters for our machine learning models in Table 5.

Table 5 Hyperparameters used for the machine learning models for to the cMSSM and pMSSM datasets

	Parameter	cMSSM	pMSSM
Supervised NN	Learning rate	0.001	0.0001245
	Hidden layers	5	10
	Nodes per layer	49	154
	Dropout	0.5	0.0
	Activation function	Sigmoid	Sigmoid
	Optimizer	ADAM	ADAM
	Batch size	128	128
	Epochs	50	50
HMC	Step size	0.025	0.008
	Number of steps	12	12
	Mass	1.0	1.0
	Chain length	5000	5000
	Burn-in steps	1000	1000
	Number of chains	100	100
NF	Num transforms	3	3
	Batch size	1024	1024
	Epochs	300	300
	B	2.0	2.0
	NN hidden features	64	64

References

1. S.P. Martin, A supersymmetry primer. Adv. Ser. Direct High Energy Phys. **21**, 1–153 (2010). arXiv:hep-ph/9709356
2. T. Cohen, J.G. Wacker, Here be dragons: the unexplored continents of the CMSSM. J. High Energy Phys. **2013**(9) (2013). https://doi.org/10.1007/JHEP09(2013)061
3. D. Ghosh, M. Guchait, S. Raychaudhuri, D. Sengupta, How constrained is the constrained MSSM?. Phys. Rev. D **86**(5) (2012). https://doi.org/10.1103/PhysRevD.86.055007
4. C. Han, K..-i Hikasa, L. Wu, J..M. Yang, Y. Zhang, Status of CMSSM in light of current LHC Run-2 and LUX data. Phys. Lett. B **769**, 470–476 (2017). https://doi.org/10.1016/j.physletb.2017.04.026
5. B.C. Allanach, Impact of CMS multi-jets and missing energy search on CMSSM fits. Phys. Rev. D **83**(9) (2011) . https://doi.org/10.1103/PhysRevD.83.095019
6. O. Buchmueller, R. Cavanaugh, A.D. Roeck, M.. J.. Dolan, J.. R. Ellis, H. Flächer, S. Heinemeyer, G. Isidori, J. Marrouche, D.. M.. Santos, K.. A. Olive, S. Rogerson, F.. J.. Ronga, K.. J.. de Vries, G. Weiglein, The CMSSM and NUHM1 after LHC Run 1. Eur. Phys. J. C **74**(6), 2922 (2014). https://doi.org/10.1140/epjc/s10052-014-2922-3
7. M. Bridges, K. Cranmer, F. Feroz, M. Hobson, R. Ruiz de Austri, R. Trotta, A coverage study of the CMSSM based on ATLAS sensitivity using fast neural networks techniques. J. High Energy Phys. **2011**(3) (2011). https://doi.org/10.1007/JHEP03(2011)012
8. M. Cahill-Rowley, J.. L. Hewett, A. Ismail, T.. G. Rizzo, Lessons and prospects from the pMSSM after LHC Run I. Phys. Rev. D **91**, 055002 (2015). https://doi.org/10.1103/PhysRevD.91.055002
9. M. Cahill-Rowley, J.L. Hewett, A. Ismail, T.G. Rizzo, pMSSM Studies at the 7, 8 and 14 TeV LHC (2013)
10. G. Aad, B. Abbott, J. Abdallah, O. Abdinov, R. Aben, M. Abolins, O.S. AbouZeid, H. Abramowicz, H. Abreu, et al. Summary of the ATLAS experiment's sensitivity to supersymmetry after LHC Run 1 – interpreted in the phenomenological MSSM. J. High Energy Phys. **2015**(10) (2015). https://doi.org/10.1007/JHEP10(2015)134
11. V. Khachatryan, A. M. Sirunyan, A. Tumasyan, W. Adam, E. Asilar, T. Bergauer, J. Brandstetter, E. Brondolin, M. Dragicevic, et al., Phenomenological MSSM interpretation of CMS searches in pp collisions at $\sqrt{s} = 7$ and 8 TeV. J. High Energy Phys. **2016**(10) (2016). https://doi.org/10.1007/JHEP10(2016)129
12. S. Caron, J.. S. Kim, K. Rolbiecki, R.. R. de Austri, B. Stienen, The BSM-AI project: SUSY-AI-generalizing LHC limits on supersymmetry with machine learning. Eur. Phys. J. C **77**(4), 257 (2017). https://doi.org/10.1140/epjc/s10052-017-4814-9
13. B. Kronheim, M. Kuchera, H. Prosper, A. Karbo, Bayesian neural networks for fast SUSY predictions. Phys. Lett. B **813**, 136041 (2021). https://doi.org/10.1016/j.physletb.2020.136041
14. A.. H. Chamseddine, R. Arnowitt, P. Nath, Locally supersymmetric grand unification. Phys. Rev. Lett. **49**, 970–974 (1982). https://doi.org/10.1103/PhysRevLett.49.970
15. MSSM Working Group Collaboration, A. Djouadi et al., The minimal supersymmetric standard model: Group summary report. In: GDR (Groupement De Recherche)—Supersymetrie, vol. 12 (1998). arXiv:hep-ph/9901246
16. G. Aad, T. Abajyan, B. Abbott, J. Abdallah, S. Abdel Khalek, A. Abdelalim, O. Abdinov, R. Aben, B. Abi, M. Abolins et al., Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. Phys. Lett. B **716**(1), 1–29 (2012). https://doi.org/10.1016/j.physletb.2012.08.020
17. S. Chatrchyan, V. Khachatryan, A. Sirunyan, A. Tumasyan, W. Adam, E. Aguilo, T. Bergauer, M. Dragicevic, J. Erö, C. Fabjan et al., Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. Phys. Lett. B **716**(1), 30–61 (2012). https://doi.org/10.1016/j.physletb.2012.08.021
18. D.. N. Spergel, L. Verde, H.. V. Peiris, E. Komatsu, M.. R. Nolta, C.. L.. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut et al., First-year Wilkinson microwave anisotropy probe (WMAP) observations: determination of cosmological parameters. Astrophys. J. Suppl. Ser. **148**(1), 175–194 (2003). https://doi.org/10.1086/377226
19. C.. L.. Bennett, D. Larson, J.. L. Weiland, N. Jarosik, G. Hinshaw, N. Odegard, K.. M.. Smith, R.. S. Hill, B. Gold, M. Halpern, E. Komatsu, M.. R. Nolta, L. Page, D.. N.. Spergel, E. Wollack, J. Dunkley, A. Kogut, M. Limon, S..S.. Meyer, G.. S. Tucker, E.. L. Wright, Nine-year Wilkinson microwave anisotropy probe

(WMAP) observations: final maps and results. Astrophys. J. Suppl. Ser. **208**(2), 20 (2013). https://doi.org/10.1088/0067-0049/208/2/20

20. G. Belanger, F. Boudjema, A. Pukhov, A. Semenov, micrOMEGAs: a tool for dark matter studies (2010)

21. B. Allanach, SOFTSUSY: a program for calculating supersymmetric spectra. Comput. Phys. Commun. **143**(3), 305–331 (2002). https://doi.org/10.1016/S0010-4655(01)00460-X

22. P. Athron, J.-H. Park, T. Steudtner, D. Stöckinger, A. Voigt, Precise Higgs mass calculations in (non-)minimal supersymmetry at both high and low scales. J. High Energy Phys. **2017**(1) (2017). https://doi.org/10.1007/JHEP01(2017)079

23. C.M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)* (Springer, Berlin, 2006)

24. D.P. Kingma, J. Ba, Adam: a method for stochastic optimization. arXiv:1412.6980 [cs.LG]

25. R.M. Neal, MCMC using Hamiltonian dynamics (2012)

26. M. Betancourt, A conceptual introduction to Hamiltonian Monte Carlo (2018)

27. G. Papamakarios, T. Pavlakou, I. Murray, Masked autoregressive flow for density estimation. arXiv:1705.07057 [stat.ML]

28. C. Durkan, A. Bekasov, I. Murray, G. Papamakarios, Neural spline flows. arXiv:1906.04032 [stat.ML]

29. J.L. Feng, K.T. Matchev, T. Moroi, Focus points and naturalness in supersymmetry. Phys. Rev. D **61**, 075005 (2000). arXiv:hep-ph/9909334