

UC Irvine

UC Irvine Electronic Theses and Dissertations

Title

Machine learning in parameter spaces beyond the Standard Model: Iterative methods for high-dimensional theory generation

Permalink

<https://escholarship.org/uc/item/2114n800>

ISBN

9798191652573

Author

Baretz, Jason Benjamin

Publication Date

2026-09-22

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Machine learning in parameter spaces beyond the Standard Model: Iterative methods for
high-dimensional theory generation

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Physics

by

Jason Baretz

Dissertation Committee:
Professor Daniel Whiteson, Chair
Professor Shirley Li
Professor Pedro Ochoa-Ricoux

DEDICATION

To George and Lynda Feldstein,
Roger and Carol Baretz,
Anne and Bill Baretz,
and my lovely wife Swati Rath-Baretz

TABLE OF CONTENTS

	Page
LIST OF FIGURES	v
LIST OF TABLES	xii
ACKNOWLEDGMENTS	xiv
VITA	xvi
ABSTRACT OF THE DISSERTATION	xviii
1 Introduction	1
2 Visualization and Efficient Generation of Constrained High-dimensional Theoretical Parameter Spaces	4
2.1 Introduction	5
2.2 Task	10
2.3 Dataset	11
2.4 Approach	13
2.4.1 Why not a VAE?	13
2.4.2 SAFESPAM	14
2.4.3 CHUNC	20
2.4.4 CHUNC2	21
2.5 Results	23
2.5.1 SAFESPAM	23
2.5.2 CHUNC and CHUNC2 for Triple Constraint	35
2.6 Discussion and Conclusion	39
2.6.1 Strengths and Weaknesses	39
2.6.2 Future Avenues	42
2.6.3 Conclusion	43
3 Iterative Flows: Towards an exhaustive search of the PMSSM	45
3.1 Introduction	45
3.2 Task & Initial Dataset	48
3.3 Approach	49

3.3.1	A Brief Introduction to Normalizing Flows	49
3.3.2	Iterating Our Flows	50
3.4	Results and Discussion	52
3.4.1	Improvements in Efficiency	57
3.4.2	Comparability to other methods	60
3.5	Conclusion	63
4	Reversing Spaces: Towards an inverse mapping of the PMSSM	65
4.1	Introduction	66
4.2	Task & Datasets	69
4.2.1	Task	69
4.2.2	Correctional Dataset	70
4.2.3	Smart Iteration Dataset	71
4.3	Approach	72
4.3.1	The Correctional Network	72
4.3.2	Iterative Corrections	73
4.3.3	Smart Iteration	74
4.4	Preliminary Results	75
4.4.1	cMSSM results	76
4.4.2	pMSSM results	78
4.5	Discussion and Conclusion	80
4.5.1	Future Avenues	80
4.5.2	Conclusion	82
	Bibliography	83
	Appendix A Hyperparameters and Supplementary Plots	96

LIST OF FIGURES

	Page
2.1 A simplified diagram of SAFESPAM for the cMSSM. GUT-scale inputs are mapped to an abstract 2D latent space, where they are shaped to a target distribution via loss function. Weak scale values that accompany each theory point are used only to classify points as valid (orange) or invalid (blue), and do not enter the network as inputs. Valid points shown in orange cluster towards the center of the latent space. The mapping f from input theory space to latent 2D space is referred to as the <i>encoder</i> and the return mapping g as the <i>decoder</i> . SAFESPAM tries to reconstruct the GUT scale parameters as accurately as possible by minimizing the L2 term in the loss function, ensuring that the outputs from the decoder match the corresponding inputs to the encoder. Sampling new points in the valid region of the latent space allows for efficient generation of new valid points in GUT space.	9
2.2 A cartoon illustration of the desired outcome of training on the latent space. Minimizing the clustering loss term pushes valid points (shown in orange) towards the latent origin while minimizing the Sliced Wasserstein loss fits those latent valid points to a target distributions. Shown in blue, invalid points (if present in training data) are pushed outward by the clustering loss term. The resulting clustering of valid points in the latent space allows for straightforward, high-efficiency sampling. Minimizing the L2 loss (not pictured) allows the network to successfully map back and forth between this structured latent space and the constrained MSSM parameter space.	17
2.3 A cartoon illustration of the KDE sampling methodology. In CHUNC and CHUNC2, step 1 is omitted and known latent valids from the training data are used instead.	19
2.4 A cartoon illustration of the bandwidth shrinking procedure in our iterative KDE sampling methodology. As the number of known valid points grows with each successive iteration, we shrink the bandwidth used to estimate the density at each point in the latent space, allowing for higher precision. . . .	19
2.5 An example of KDE sampling in the pMSSM: An initial uniform scan of the latent space (left) demarcates the shape of the valid core, and the valid points found in that scan are utilized via kernel density estimation to sample the core efficiently (right). White spaces represent areas that are avoided by KDE to improve efficiency.	20

2.6	Distribution of sampled points in the two-dimensional latent space, for the case of learning to generate cMSSM points which satisfy the Higgs mass constraint. Points which satisfy (fail) the Higgs mass constraint are marked Valid (Invalid). Invalid regions in the center are partially avoided due to KDE sampling.	26
2.7	Distribution of SAFESPAM-generated points (red) in the cMSSM in each of the five cMSSM parameters, under the Higgs mass constraint, compared to the distribution of the training dataset (blue).	26
2.8	Distribution of sampled points in the two-dimensional latent space, for the case of learning to generate pMSSM points which satisfy the Higgs mass constraint. Points which satisfy (fail) the Higgs mass constraint are marked Valid (Invalid). Shown are points sampled via KDE, where invalid regions in the center are partially avoided due to KDE sampling. Trial 1 is shown on the left and Trial 2 on the right, showcasing the sensitivity of the latent space structure to the hyperparameters.	27
2.9	Distribution of SAFESPAM-generated points (red for Trial 1, orange for Trial 2) in the pMSSM in each of the nineteen pMSSM parameters, compared to the distribution of the training dataset (blue). While both trials achieve strong efficiency, they sample different portions of the parameter space.	28
2.10	Distributions in the two-dimensional latent space for the cMSSM dark matter study, where points which satisfy (fail) the dark matter constraint are marked Valid (Invalid). Left shows the distribution of valid points in the training data, with a well-formed core. However, many new points sampled from this core do not satisfy the dark matter constraint (right), perhaps due to insufficient resolution of DM-valid features in the training data. Successive KDE iterations allowed for precise, efficient sampling in the remaining regions of high validity (see Figs 2.11 and 2.13).	29
2.11	Distributions of sampled points in the two-dimensional latent space in four successive rounds of sampling for the cMSSM dark matter case. Points which satisfy (fail) the dark matter constraint are marked Valid (Invalid). In each round, newly discovered valid points are used to update the kernel density estimation of valid regions of the latent space, which is used to perform the next round of sampling. As a result, the ‘crust’ of invalid points shown in blue shrinks away with each successive iteration. The efficiencies for these 4 rounds are 0.724, 0.887, 0.945, and 0.972 respectively.	30
2.12	Distribution of SAFESPAM-generated points in the cMSSM (red) in each of the five cMSSM parameters under the dark matter constraint, compared to the distribution of the training dataset (blue).	31

2.13	Distributions in the two-dimensional latent space for the pMSSM dark matter study, where points which satisfy (fail) the dark matter constraint are marked Valid (Invalid). Shown are distributions of sampled points in successive rounds, where in each round valid points from previous rounds are used to perform a kernel density estimation of valid regions of the latent space for the next round of sampling. The initial scan in the top left frame samples via uniform distribution over a region determined by the position of latent valid training data. The efficiency for each round is 0.049, 0.432, 0.881, 0.950, 0.979, respectively.	31
2.14	Distribution of SAFESPAM-generated points in the pMSSM (red) in each of the nineteen pMSSM parameters under the dark matter constraint, compared to the distribution of the training dataset (blue).	32
2.15	Distributions in the two-dimensional latent space for the cMSSM study, where points which satisfy (fail) the Higgs mass constraint are marked Valid (Invalid), and valid points which also satisfy the dark matter constraint are marked as Double-Valid. In the left, the initial scan identifies a core of valid Higgs points. On the right are the newly generated points which satisfy both constraints.	33
2.16	Distributions in the two-dimensional latent space for the pMSSM study, where points which satisfy (fail) the Higgs mass constraint are marked Valid (Invalid), and valid points which also satisfy the dark matter constraint are marked as Double-Valid. In the top left, the initial scan identifies a core of valid Higgs points. In the top right, the subset of those points which also satisfy the dark matter constraints are identified. In the bottom left are points generated using a KDE built from the double-valid points, and in the bottom right are the new points which satisfy both constraints.	34
2.17	Distribution of CHUNC (red) and CHUNC2 generated (orange) points in each of the five cMSSM parameters under the triple constraint, compared to the distribution of the training dataset (blue).	38
2.18	Distribution of CHUNC (red) and CHUNC2 generated (orange) points in each of the nineteen pMSSM parameters under the triple constraint, compared to the distribution of the training dataset (blue).	39
3.1	A simplified 2D cartoon demonstrating the issue with applying multiple experimental constraints at once on the same dataset/scan. Insufficient resolution means that irregular but desirable regions in theory space may be overlooked. Interesting and unique theories may hide in these enclaves.	48
3.2	A simplified cartoon demonstrating an iterative training regimen. While our methodology uses the cumulative valid points from all prior datasets to train the next flow, in practice the majority of these points frequently come from the most recent preceding flow.	51
3.3	The normalized distributions of Higgs-valid pMSSM subsets are shown here: S_0^1 from the naive scan D_0 in red and S_1^1 from the flow-generated D_1 in blue. The proportionalities are a close match, demonstrating that our flow is behaving as expected in capturing the proportionality of the Higgs-valid subspace.	52

3.4	The normalized distributions of '2-valid' pMSSM subsets are shown here: S_0^2 from the naive scan D_0 in red, S_1^2 from D_1 in blue, and S_2^2 from D_2 in green, whose flow was trained on $S_0^2 \cup S_1^2$. We observe only minor deviations from the expected proportionality.	53
3.5	The distribution of 7-valid pMSSM points S_0^7 from the naive scan D_0 are shown in red, against the 7-valid points S_7^7 from the flow-generated D_7 which are shown in blue. Several regions where the initial scan found few or no 7-valid points are in fact still viable. See discussion below for specific ranges.	54
3.6	The weak-scale observables of the two 7-valid subsets shown in figure 3.5 are shown here, as distributions in 'constraint space'. S_0^7 from D_0 is shown in red, and S_7^7 from D_7 in blue. Note that since constraint 2 ($m_{LSP} = \chi_1^0$) is not a range of values, the actual value of m_{LSP} for points which meet this constraint is plotted here instead.	55
3.7	Selected 2D histograms comparing the 7-valid distributions S_0^7 from D_0 in red, and S_7^7 from D_7 in blue. Bins containing only a single theory point from S_7^7 are shown in charcoal for improved visibility in sparse regions and the axes for GUT-scale parameters are set to exclude points outside the initial scan boundaries. Note the islands/peninsulas found around $600 \text{ GeV} \lesssim m_{LSP} \lesssim 1100 \text{ GeV}$ which the naive scan largely missed, but which the flow was able to sample admirably. The parameters plotted are as follows: a.) a_μ^{SUSY} vs. m_{LSP} b.) $\frac{BR(B_u \rightarrow \tau \nu)}{BR(B_u \rightarrow \tau \nu)_{SM}}$ vs. m_{LSP} c.) m_{LSP} vs. M_1 d.) m_{LSP} vs. M_2 e.) m_{LSP} vs. M_A f.) m_{LSP} vs. $\tan \beta$	58
3.8	Selected 2D histograms comparing the 7-valid distributions S_0^7 from D_0 in red, and S_7^7 from D_7 in blue. Bins containing only a single theory point from S_7^7 are shown in charcoal for improved visibility in sparse regions, and the axes for GUT-scale parameters are set to exclude points outside the initial scan boundaries. Note the islands/peninsulas found around $m_{\tilde{L}_1} \gtrsim 1500 \text{ GeV}$ which the naive scan largely missed, but which the flow was able to sample admirably. The parameters plotted are as follows: a.) a_μ^{SUSY} vs. $m_{\tilde{L}_1}$ b.) m_{LSP} vs. $m_{\tilde{L}_1}$ c.) $m_{\tilde{e}_1}$ vs. $m_{\tilde{L}_1}$	59
3.9	The distribution of distances from each point in S_0^7 in red and S_7^7 in blue to the nearest non-identical neighboring point in S_0^7 . Distances were calculated with axes normalized using the values in Table 3.2. The ranges of these distributions suggest our flow was able to discover new 7-valid points that the naive scan missed due to low resolution (i.e. distances $\lesssim 1.5$) as well as those outside the range of the initial scan (See Appendix A.3.1)	61

3.10	A scatterplot of recent SOTA pMSSM search methods showing efficiency vs. number of experimental constraints. The studies cited are as follows: CMA-ES [44], NSGAI [44], TPE [44], HMC [77], Normalizing Flows [77], SBI [104], NPE (amortized SBI) [38] and Genetic Algorithms [3] are plotted against our results (listed here as "Iterative Flows"). Many of the proportional search papers listed here also conducted single-constraint trials (typically using the Higgs Mass constraint) which achieved higher efficiencies than their respective two-constraint runs. These single-constraint runs are omitted here for legibility. Please note the following caveats and stipulations: * Efficiency values for [44] are taken from [<i>Phys. Rev. D</i> 107, 035004 (2023)], which differ from the values specified in the arxiv preprint. † Efficiency values for [77] are taken from [<i>Eur. Phys. J. C</i> 81, 1138 (2021)] which differ from the values specified in the arxiv preprint. The evaluation software used in [77] has known computational differences which may have led to artificially lowered efficiencies, as noted in Footnote 10 of [44]. ‡ The convention used in [3] contains 'aggregated composite' constraints such as HiggsSignals [15] which combines multiple signals into a single experimental constraint with a single χ^2 contribution. We count each such composite as one individual constraint, and as such, when observables which appear within that composite block are instead studied individually, as in [38], we group and count those as one individual constraint as well. This discrepancy in how individual constraints are denoted is illustrative of the broader difficulties in comparing results across multiple pMSSM search studies. We also note that [3] contains two separate trials with nearly identical constraints, with the exception that δa_μ^{SUSY} in trial 1 is replaced with the GCE constraint [35] in trial 2 (in both cases, no theory points which meet all 13 constraints were found).	62
4.1	A simple cartoon in a 2D slice of theory space demonstrating a desired correction where the vector ΔT maps the proposed T_{old} to the nearest point with the desired weak signature (here dark matter relic density).	69
4.2	A series of diagrams representing the procedures for a single correction, iterative correction, and 'smart iteration' correction, respectively	72
4.3	The average length of the correction vector $ \delta $ is plotted over 100 iterative corrections for a batch of 10k cMSSM points. We see approximately asymptotic behavior, suggesting that the network is getting closer to the desired solution with each iteration, not vacillating wildly between many possible options.	74

4.4	"Input/Output" plots show the results of 40x iterative corrections in the cMSSM. Standard iterative corrections are shown on the left, and 'smart iteration' on the right. The x-axis shows the input weak value W while the y-axis shows the true weak signature of the generated GUT point, calculated using SOFTSUSY and MicrOMEGAS. A 'perfect' correction would sit exactly on the diagonal $y=x$, and in practice we see a significant number of points cluster along this trend line, though this is much more prominent for the higgs constraint than for dark matter. We also note that smart iteration shows a visible improvement in dark matter performance over standard iteration, as expected.	75
4.5	cMSSM points generated via smart iteration as seen in 4.4 are plotted in 2D slices of the cMSSM parameter space and shaded by the log of the dark matter error $Log((\Omega_{DM}h^2)_{Desired} - (\Omega_{DM}h^2)_{True})$. We observe that the model performs best in areas that correspond to high LOKV density, and worse in sparser regions.	77
4.7	The total number of cMSSM points accepted via smart iteration is plotted as a function of the number of iterations. We see diminishing returns at around 41.5% of the 10k sample. The remaining $\mathcal{O}(5.85k)$ points are taken as-is after 40 iterations.	77
4.6	cMSSM points generated via smart iteration as seen in 4.4 are plotted in weak space and shaded by the log of the dark matter error $Log((\Omega_{DM}h^2)_{Desired} - (\Omega_{DM}h^2)_{True})$. We observe that the model performs best in areas that correspond to high LOKV density, and worse in sparser regions.	78
4.8	"Input/Output" plots show the results of a single correction in the pMSSM, in both standard and log scale. The x-axis shows the input weak value W while the y-axis shows the true weak signature of the generated GUT point, calculated using SOFTSUSY and MicrOMEGAS. A 'perfect' correction would sit exactly on the diagonal $y=x$, and while we see significant clustering along this trendline for the Higgs constraint, the dark matter prediction has significant inaccuracy for most generated points. Improving this is the primary remaining task for this project.	79

- 4.9 Histograms show the error across both constraints for two separate 10k runs: a 40x smart iteration run in the cMSSM on the left and a single pMSSM correction on the right. We note the following: 1.) A dotted line at $|(\Omega_{DM}h^2)_{Desired} - (\Omega_{DM}h^2)_{True}| = 0.01$ denotes our desired minimum accuracy. 2.) The dark matter predictability in the cMSSM case significantly outperforms that of the pMSSM, though both perform well on the Higgs constraint 3.) An artifact exists around $|(\Omega_{DM}h^2)_{Desired} - (\Omega_{DM}h^2)_{True}| = \mathcal{O}(1.0)$ and $|(m_h)_{Desired} - (m_h)_{True}| = \mathcal{O}(126.0)$ GeV which results from points that are 'theoretically invalid' (i.e. regions of parameter space where weak signatures are not defined/calculable). Our evaluation software marks these as "-1", creating artificial spikes corresponding to $(\Omega_{DM}h^2)_{Desired} + 1$ and $(m_h)_{Desired} + 1$. 4.) We note that the pMSSM run generated significantly fewer theoretical invalids than the cMSSM run despite using only a single correction. It is possible that theoretically invalid points are simply less common in the pMSSM due to the increase in number of degrees of freedom, or alternatively, perhaps theoretical invalids are more densely packed in regions that adequately satisfy the DM constraint. 80
- 4.10 A simplified cartoon of the proposed embedding correctional network: 1.) The initial 'bad guess' points are mapped to the embedding space via an encoder. 2.)The correctional network performs the nudge strictly in latent space to yield the desired weak signature 3.)The corrected points are mapped back to GUT-space via decoder. 82

LIST OF TABLES

	Page
2.1 Range for cMSSM parameters used in the initial brute-force generation of the training dataset.	11
2.2 Range for pMSSM parameters used in the initial brute-force generation of the training dataset.. An additional pre-processing step was taken to manually remove the gaps in $ M_1 < 0.05$ TeV, $ M_2 < 0.1$ TeV, and $ \mu < 0.1$ TeV before training.	12
2.3 Comparison of the efficiency of generation of experimentally-valid points in naive scanning versus SAFESPAM, for several experimental constraints (Higgs mass m_H , dark matter relic density Ω_{DM} , or both) and two theoretical spaces (5-dim cMSSM, 19-dim pMSSM).	35
2.4 Comparison of the Core-Forming and Incompleteness Metrics for various SAFESPAM trials for several experimental constraints (Higgs mass m_H , dark matter relic density Ω_{DM}) and two theoretical spaces (5-dim cMSSM, 19-dim pMSSM).	35
2.5 Comparison of the efficiency of generation of experimentally-valid points in naive scanning versus CHUNC and CHUNC2 methods for three experimental constraints (Higgs mass m_H , dark matter relic density Ω_{DM} , and the lightest supersymmetric particle (LSP) being a neutralino) and two theoretical spaces (5-dim cMSSM, 19-dim pMSSM). The uncertainty across all trials is $\sim \mathcal{O}(0.01)$, estimated by running 10 independent samplings for each case.	37
2.6 Comparison of the Core-Forming and incompleteness Metrics for CHUNC and CHUNC2 for the 'triple' experimental constraint (Higgs mass $m_H \cap$ dark matter relic density Ω_{DM} <i>cap</i> Neutralino LSP,) in two theoretical spaces (5-dim cMSSM, 19-dim pMSSM). Note that CHUNC2's categorical variable is excluded from calculations of the Core-Forming metric.	37
3.1 A list of relevant experimental constraints for our search, presented in order of their application.	49
3.2 Ranges for the pMSSM parameters used in the initial brute-force scan D_0 . Before training an additional pre-processing step was taken to manually remove the gaps in $ M_1 < 0.05$ TeV, $ M_2 < 0.1$ TeV, and $ \mu < 0.1$ TeV. This step was reversed during the generation of new data.	50

3.3	Table of efficiencies for each flow and set of constraints, where a "valid point" in each row is defined as one that meets both a.) the specified constraint, and b.) all prior constraints. Each dataset consists of $\mathcal{O}(22M)$ theory points with the exception of D_0 which contains $\mathcal{O}(20.9M)$ points. The efficiency term for the i th leg of the training under j cumulative constraints, following the ordering in table 3.1, can be written as $E_i^j = \frac{ S_i^j }{ D_i }$. The $E_i^{j=i}$ terms are marked here in bold.	56
-----	---	----

ACKNOWLEDGMENTS

I am deeply grateful for my advisor Daniel Whiteson, whose guidance and support has been essential to my completion of this degree.

I would also like to thank the other members of my defense committee: Shirley Li and Pedro Ochao Ricoux.

Significant thanks and recognition are owed to Jacob Hollingsworth for his mentorship over the course of many conversations, as well as to my co-author Nicholas Carrara.

I would be remiss to not also include thanks for my friends in the cohort: Nicholas Duong, Dylan Green, Mark Hayward and Adreja Mondol, without whom I would have left the house significantly less over the past 7 years.

Acknowledgments for Chapter 2:

The text of chapter 2 of this dissertation is a reprint of the material as it appears in J. Baretz, N. Carrara, J. Hollingsworth, and D. Whiteson, "Visualization and efficient generation of constrained high-dimensional theoretical parameter spaces," *Journal of High Energy Physics*, vol. 2023, no. 11, p. 62, 2023, used with permission from Springer (Creative Commons Attribution 4.0 International License). The coauthors listed in this publication are Jason Baretz, Nicholas Carrara, Jacob Hollingsworth, and Daniel Whiteson. Daniel Whiteson directed and supervised research which forms the basis for the dissertation.

Special thanks to Aishik Ghosh and Jessica N. Howard for their valuable advice and discussion, as well as Max Fieg, Kevin Greif, Gregor Kasieczka, Michael Ratz, and David Shih for their feedback and perspective.

This material is based upon work supported by the Department of Energy National Nuclear Security Administration through the Nuclear Science and Security Consortium under Award Number DE-NA0003996. **Disclaimer:** This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or limited, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The view and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Acknowledgments for Chapter 3: This work was supported by the DOE, High Energy Physics Research and Technology Division under grant DE-SC0023966. Training was

conducted in part using the Greenplanet cluster (NSF Grant No. CHE- 0840513) and for evaluations utilized the infrastructure for high-performance and high-throughput computing, research data storage and analysis, and scientific software tool integration built, operated, and updated by the Research Cyberinfrastructure Center (RCIC) at the University of California, Irvine (UCI). The RCIC provides cluster-based systems, application software, and scalable storage to directly support the UCI research community. <https://rcic.uci.edu>. Special thanks to Jake Rudolph for their feedback and perspective.

Acknowledgments for Chapter 4: This work was supported by the DOE, High Energy Physics Research and Technology Division under grant DE-SC0023966. Training was conducted in part using the Green Planet cluster (NSF Grant No. CHE- 0840513) and initial dataset evaluations (for LOKV) utilized the infrastructure for high-performance and high-throughput computing, research data storage and analysis, and scientific software tool integration built, operated, and updated by the Research Cyberinfrastructure Center (RCIC) at the University of California, Irvine (UCI). The RCIC provides cluster-based systems, application software, and scalable storage to directly support the UCI research community. <https://rcic.uci.edu>. We are grateful to Nicholas Carrara, Aishik Ghosh, Junze Liu, Gwendolyn Romanelli and Alexander Shmakov for their valuable advice and discussion throughout various portions of this project.

VITA

Jason Baretz

EDUCATION

Doctor of Philosophy in Physics and Astronomy

University of California, Irvine

2026

Irvine, CA

Bachelor of Arts in Physics

Cornell University, College of Arts and Sciences

2019

Ithaca, NY

RESEARCH EXPERIENCE

Graduate Research Assistant

University of California, Irvine

2019–2026

Irvine, California

TEACHING EXPERIENCE

Teaching Assistant

University of California, Irvine

2019–2023

Irvine, California

REFEREED JOURNAL PUBLICATIONS

Towards AI-assisted neutrino flavor theory design **2026**

Jason Baretz, Max Fieg, Vijay Ganesh, Aishik Ghosh, V. Knapp-Perez, Jake Rudolph,
Daniel Whiteson

Communications Physics

**Visualization and efficient generation of constrained
high-dimensional theoretical parameter spaces** **2023**

Jason Baretz, Nicholas Carrara, Jacob Hollingsworth and Daniel Whiteson

Journal Of High Energy Physics

**On The Calculation Of Invariant Tensors In Gauge The-
ories** **2023**

Yahya Almumin, *Jason Baretz*, Eric Bryan, Arvind Rajaraman

Acta Physica Polonica B

REFEREED CONFERENCE PUBLICATIONS

**Towards The Optimization Of Photocathode Proper-
ties Via Surface Science Techniques: A Study On Cs3sb** **2019**

Thin Film Growth

Alice Galdi, Jan Balajka, Ivan Bazarov, Luca Cultrera, William DeBenedetti, Melissa
Hines, Frank Ikponmwon, Jared Maxson, Sean McBride, *Jason Baretz* et al.

Proceedings of the North American Particle Accelerator Conference

ABSTRACT OF THE DISSERTATION

Machine learning in parameter spaces beyond the Standard Model: Iterative methods for high-dimensional theory generation

By

Jason Baretz

Doctor of Philosophy in Physics

University of California, Irvine, 2026

Professor Daniel Whiteson, Chair

Generative models with iterative methodologies are used to probe high-dimensional subspaces of the Minimal Super Symmetric Model (MSSM) in a variety of contexts. First, a dimensional reduction framework is used to sample theory points from a low-dimensional latent space and then de-code them back into theory parameters. Iterative density estimation sampling is used here to maximize efficiency as well as to filter by additional constraints post facto without further training. Next, an iterative usage of generative flows is used to produce models which can sample theory points under multiple simultaneous, without loss of search resolution. An initial flow is trained on a single constraint, after which a new dataset generated from that flow is used to create training data for the learning of two simultaneous constraints. This process is repeated up to seven simultaneous constraints, with high efficiency. Finally, we explore a new methodology which approximates an inverse mapping from experimental observable space to theory parameter space by learning to make small corrections to a 'bad guess' candidate theory. Iterative corrections, taking the previous output theory point as a new input, are used to achieve sub-constraint-window accuracy. A secondary network is also trained to monitor the progress of these iterations in order to maximize efficacy.

Chapter 1

Introduction

In a broad sense, the process of deep machine learning can be thought of naively as a transfer of information: patterns contained within training data is transferred into the form of network weights and biases via algorithms such as back-propagation [124, 118]. The ability for such neural networks to extrapolate this way is often explicated via the manifold hypothesis, the idea that the useful information needed for learning actually exists along a lower dimensional manifold representation that sits within complex high-dimensional datasets [123, 129]. However, we also know in general that excessive dimensional reduction can result directly in information loss [125, 78, 131, 43]. Another potential bottleneck is that in most training contexts, training data is considered a finite resource. So, especially in generative applications, a relevant question might be: how might one 'update' a sampling method in light of new information. Bayesian inference gives a solid mathematical and statistical underpinning to this sort of question [62, 21, 10], but practically speaking how does one improve a faulty model in cases where that model's inference is not strictly Bayesian?

Iterative methodologies offer an appealing answer to this question: using an existing model's output data as a springboard allows us to inject new information and improve a model's

generative outcomes, sometimes without even requiring additional training.

In the scope of theoretical high-energy physics (HEP), generative ML tasks often consist of searching through a high-dimensional parameter space to find individual theories which agree with experiment. This is often a non-trivial task both due to the high-dimensionality of the space and the complex non-linear relationship between theory parameters and experimental observables. While there are many methodologies equipped for this task [77, 104, 4, 44, 46, 3, 32, 96, 23, 112, 22, 97, 98, 14, 13], how can we be sure such methods are maximizing their utility in a landscape where brute-force dataset augmentation is often time and resource expensive? This thesis focuses on answering the question "how can iterative methods be utilized to improve the outcomes of existing generative methodologies in HEP applications beyond the standard model?".

In chapter 2 we train a sliced Wasserstein auto encoder to map GUT theory points to a 2D latent space where sampling and visualization are significantly easier. Crucially, our mapping is contrastive, such that experimentally valid points cluster together into a 'core'. We utilize an iterative formulation of Kernel Density Estimation to achieve high efficiency in sampling this core in both the cMSSM pMSSM. Because our mapping is deterministic, we can also use iterative sampling to probe sub-regions of the valid core which meet additional constraints, despite the model not being trained on those constraints.

In chapter 3 we conduct an in-depth probing of the pMSSM utilizing Normalizing Flows in an iterative context to prevent loss of search resolution. We train first on a single experimental constraint, then use the resulting model to generate new training data, which we use learn two constraints at once, and so forth. In all, we are able to probe seven simultaneous constraints resulting in an improvement in efficiency of several orders of magnitude.

In chapter 4 we introduce the framework for a 'correctional network' which is designed to approximate an inverse mapping of GUT-to-Weak evaluation software to take advantage of

the tractable properties of the weak-scale observable space when sampling. Our network functions by learning to correct a 'bad guess' in GUT space such that the resulting theory point has the specified weak signature. We utilize the network output to then perform iterative corrections, treating the 'corrected point' as a new input point. We find that the average distance of our corrections seems to converge towards zero as we increase the number of iterations, suggesting that the model is indeed closing in on a specific answer as opposed to fluctuating between possible options. We are able to reach sub-constraint-window levels of precision in 'reversing' two experimental constraints, though the results shown here are preliminary. We detail what work remains for this project in the discussion section.

Chapter 2

Visualization and Efficient Generation of Constrained High-dimensional Theoretical Parameter Spaces

The text of this chapter of the dissertation is a reprint of the material [13] as it appears in J. Baretz, N. Carrara, J. Hollingsworth, and D. Whiteson, "Visualization and efficient generation of constrained high-dimensional theoretical parameter spaces," Journal of High Energy Physics, vol. 2023, no. 11, p. 62, 2023, used with permission from Springer Creative Commons Attribution 4.0 International License. The coauthors listed in this publication are Nicholas Carrara, Jacob Hollingsworth, and Daniel Whiteson. Daniel Whiteson directed and supervised research which forms the basis for the dissertation.

Abstract— We describe a set of novel methods for efficiently sampling high-dimensional parameter spaces of physical theories defined at high energies, but constrained by experimental measurements made at lower energies. Often, theoretical models such as supersymmetry are defined by many parameters, $\mathcal{O}(10 - 100)$, expressed at high energies, while relevant experimental constraints are often defined at much lower energies,

preventing them from directly ruling out portions of the space. Instead, the low-energy constraints define a complex, potentially non-contiguous subspace of the theory parameters. Naive scanning of the theory space for points which satisfy the low-energy constraints is hopelessly inefficient due to the high dimensionality, and the inverse problem is considered intractable. As a result, many theoretical spaces remain under-explored. We introduce a class of modified generative autoencoders, which attack this problem by mapping the high-dimensional parameter space to a structured low-dimensional latent space, allowing for easy visualization and efficient generation of theory points which satisfy experimental constraints. An extension without dimensional compression, which focuses on limiting potential information loss, is also introduced.

2.1 Introduction

Decades of searches for extensions to the Standard Model such as supersymmetry (SUSY) have come up empty, and future higher-energy colliders may be required in order to discover new particles and interactions. But given the vast expense and wait time incurred by such projects, it is worth exploring diligently whether the parameter spaces of these theories have been thoroughly exhausted in current datasets. Could there still be novel unexplored islands—unexcluded regions of parameter space with signatures accessible to the LHC and astroparticle experiments?

The full parameter space of the minimal super-symmetric model (MSSM) has more than 100 parameters [100], making a complete exploration daunting. Out of desperate pragmatism, the search space is often restricted to a more tractable subset of parameters based on theoretical inclinations. Common examples include the $(4 + 1)$ -dimensional cMSSM as well as the 19-dimensional pMSSM. These high-dimensional spaces guide the LHC program [63, 73, 9, 25, 24, 28, 30, 2, 42], but direct experimental searches are typically performed in only one or two dimensions, with all other parameters fixed to theoretically preferred values.

The resulting surfaces of LHC exclusion occupy a vanishingly small fraction of the volume of the larger theoretical space, most of which is essentially unexamined. Each point in such a space defines a spectrum of particles which determine well-known quantities such as the Higgs boson mass and the dark matter relic density. But what portions of the full SUSY space are consistent with these weak-scale constraints and LHC exclusion results? Are there unexplored islands where promising models abound? Nobody knows.

A critical obstacle is that weak-scale constraints such as the Higgs mass or the dark matter relic density are not expressed directly in terms of the fundamental parameters of the theory, such as SUSY particle masses, instead requiring intensive calculations. This prevents us from using experimental constraints to immediately rule out large portions of the model’s parameter space or reduce its dimensionality. Instead, the weak-scale constraints define a complex, potentially non-contiguous subspace of the theory parameters. And the inverse problem, determining a map from weak-scale observables to fundamental parameters, is intractable. Instead, we are forced to scan the high-dimensional parameter space, evaluating each point for weak-scale consistency. Efforts to speed up the calculation of the weak-scale quantities with machine learning [32, 96, 23] do not solve the core problem, which is dominated by the dependence of the scanning cost on the dimensionality rather than the per-sample computation time.

Even the reduced spaces such as the pMSSM are difficult to scan using a brute-force search, so much so that the discovery of new islands of experimental consistency in this subspace relies largely on the inspiration of particle theorists intuiting their existence. Each island (e.g. “well-tempered SUSY” [40], “focus point SUSY” [57], “Natural SUSY” [111], “Gauge Mediated SUSY Breaking” [47], “General Gauge Mediation” [103]) is then carefully studied [95, 71, 79, 58, 94, 54, 26, 27, 49, 48], and novel signatures can motivate new experimental analyses. But the danger of this exploration by intuition is that one may miss viable regions. Currently, many particle physics theories include vast, mostly-unexplored parameter spaces, which may

contain undiscovered islands consistent with experimental constraints. Some of these islands may reveal new unanticipated signatures accessible at the LHC.

Fortunately, an arsenal of techniques in Artificial Intelligence (AI), and in particular statistical techniques such as Machine Learning (ML), have shown great promise at exploring and summarizing information from high-dimensional data spaces. Recent approaches for exploring high dimensional HEP spaces fall largely into two broad categories.

A class of generative method such as GANs [68], HMC [19, 107, 11], Normalizing Flows [88], and Genetic Algorithms [66], take advantage of the ability of ML to perform high-dimensional interpolation, making notable improvements in the efficiency of sampling high-dimensional subspaces consistent with experimental constraints [77, 104, 4, 44, 46, 3]. But these methods often act as black boxes, making interpretability difficult, and ameliorate but do not fundamentally solve the underlying difficulty of searching high-dimensional spaces. The increasingly high-dimensionality of SUSY-related theoretical spaces may instead require methods which map the problem to a lower-dimensional space where extensive sampling is more feasible.

The other broad category of approaches, observational dimensional reduction methods (e.g. UMAP [101], TSNE [126], non-generative auto-encoders [6], and self-organizing maps [90, 89]), analyze high dimensional datasets by mapping them to a lower-dimensional latent space [105, 76], allowing for *visualization* of any resulting structure, such as fertile islands of desirable points. For example, in the context of SUSY parameter space searches, these techniques allow for visual examination of the latent space in the hopes of finding clusters of points consistent with experimental constraints. While the goal of collapsing the high-dimensional structure into a lower dimensional space has solid mathematical underpinnings [56], actually creating useful structure is challenging, as many of the methods are unsupervised and thus cannot explicitly encourage such structure in the latent space. Often, the resulting latent space is too poorly structured for efficient sampling [92]. In other cases,

the lack of a return mapping from the latent space prevents sampling altogether. However, when used in the context of generative methods, reducing the dimensional complexity of the sampling space has the potential to directly remove the primary difficulty posed by searching high-dimensional spaces.

Additionally, Manifold Learning methods [112, 22, 97, 98] spanning both of these categories have sought to learn and sample from low-dimensional manifolds on which interesting points sit in high dimensional spaces, although they can be non-trivial and computationally expensive to train [112, 22].

In this paper, we build a new class of generative methods which map the problem to a lower-dimensional space, but specifically encourage useful structure by treating the problem as a supervised learning task. We use a modified Sliced Wasserstein Autoencoder (SWAE) [92] structure to create bidirectional maps between the GUT-scale parameter space and a *structured low-dimensional latent space*. The structure of the latent space is specialized to force valid points to cluster near the latent origin, allowing for efficient generation of new points which can be mapped back to the GUT-scale parameter space. Because the latent space is low-dimensional, it is inherently easier to search. Additionally, the added benefit of visualization due to dimensional reduction makes our maps easy to understand and interact with, and in some cases allows the addition of secondary constraints *in media res* with no additional training. Because our method reduces the complexity of the sampling space, it may be scalable in the future to tackle larger, less-constrained subsections of the MSSM. The structure of our approach shown in Fig. 2.1. To explore whether encouraging structure in the latent space is sufficient, without insisting on dimensional reduction, we also introduce a complementary formulation of our method which operates without any dimensional reduction.

We call our primary method the SWAE Algorithm For Experimentally Sound Point-Production And Mapping (SAFESPAM), and two variants without dimensional reduction we refer to

as Constraint-driven High-dimensional UNcompressed (Categorical) Clustering, or CHUNC and CHUNC2¹ [33].

The paper is outlined as follows: In Section II we detail the task and nature of our experimental constraints. Section III describes the generation and parameters of our dataset. In Section IV, we discuss our approach for SAFESPAM, CHUNC and CHUNC2, as well as our sampling procedures. Section V presents the results of our trials for various sets of experimental constraints, and Section VI contains further discussion and conclusions.

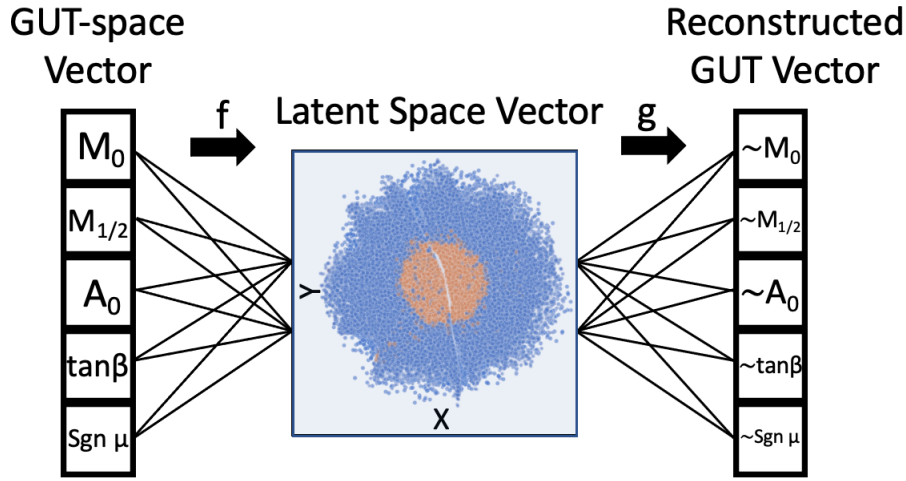


Figure 2.1: A simplified diagram of SAFESPAM for the cMSSM. GUT-scale inputs are mapped to an abstract 2D latent space, where they are shaped to a target distribution via loss function. Weak scale values that accompany each theory point are used only to classify points as valid (orange) or invalid (blue), and do not enter the network as inputs. Valid points shown in orange cluster towards the center of the latent space. The mapping f from input theory space to latent 2D space is referred to as the *encoder* and the return mapping g as the *decoder*. SAFESPAM tries to reconstruct the GUT scale parameters as accurately as possible by minimizing the L2 term in the loss function, ensuring that the outputs from the decoder match the corresponding inputs to the encoder. Sampling new points in the valid region of the latent space allows for efficient generation of new valid points in GUT space.

¹<https://github.com/infophysics/CHUNCNet>.

2.2 Task

To efficiently generate new points in high-dimensional theoretical parameter spaces which satisfy the experimental constraints placed on weak-scale observables, we seek maps between the theory space and a structured latent space which are:

- *generative*: capable of generating new points with high efficiency compared with naive brute force sampling
- *contrastive*: with structured latent spaces where valid and invalid points are separated into distinct regions
- *low-dimensional*: for ease of sampling and visualization of the latent space

In this paper, we study the performance of the SAFESPAM method in two constrained SUSY spaces, the cMSSM in 4+1 dimensions and pMSSM in 19 dimensions [41, 51, 36]. We impose two experimental constraints, the Higgs boson mass and the dark matter relic density:

- Higgs Boson Mass: $122.09 \text{ GeV} \leq m_h \leq 128.09 \text{ GeV}$ [1, 37]
- Dark Matter Relic Density: $0.08 \leq \Omega_{DM} h^2 \leq 0.14$ [122, 17]

Points are denoted as ‘valid’ whose experimental values fall within these constraints. While in principle our methods could also use other experimental variables as constraints, prior work has often focused on these [77, 104, 44]. With CHUNC and CHUNC2, we introduce a third constraint that the Lightest Supersymmetric Particle (LSP) be a neutralino. All weak scale constraints are calculated using softsusy 4.1.10 and micromegas 5.2.6 [8] [16].

2.3 Dataset

We utilize initial brute-force scans, $\mathcal{O}(10\text{-}20\text{M})$ points, of the parameters of pMSSM and cMSSM at the GUT scale to create our datasets [41, 28]. Below we denote the parameter ranges for the cMSSM (Table 2.1) and pMSSM (Table 3.2). All parameters are sampled uniformly over the specified range, with the exception of A_0 , where $\frac{A_0}{m_0}$ is sampled uniformly [77].

GUT-scale points are considered to be *valid* if they are consistent with experimental constraints, calculated via softsusy 4.1.10 and micromegas 5.2.6 [8] [16].

Table 2.1: Range for cMSSM parameters used in the initial brute-force generation of the training dataset.

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
$\text{sgn}(\mu)$	1 or -1	Sign of Higgsino mass

Table 2.2: Range for pMSSM parameters used in the initial brute-force generation of the training dataset.. An additional pre-processing step was taken to manually remove the gaps in $|M_1| < 0.05$ TeV, $|M_2| < 0.1$ TeV, and $|\mu| < 0.1$ TeV before training.

Parameter	Domain	Description
$ M_1 $	[0.05,4] TeV	Bino Mass
$ M_2 $	[0.1,4] TeV	Wino Mass
M_3	[0.4,4] TeV	Gluino Mass
$ \mu $	[0.1,4] TeV	Bilinear Higgs Mass
M_A	[0.1,4] TeV	Pseudo-scalar Higgs Mass
$ A_b $	[0,4] TeV	Trilinear Bottom Coupling
$ A_t $	[0,4] TeV	Trilinear Top Coupling
$ A_\tau $	[0,4] TeV	Trilinear τ Coupling
$m_{\tilde{L}_1}$	[0.1,4] TeV	1st Generation L.h. Slepton Mass
$m_{\tilde{L}_3}$	[0.1,4] TeV	3rd Generation L.h. Slepton Mass
$m_{\tilde{e}_1}$	[0.1,4] TeV	1st Generation R.h. Slepton Mass
$m_{\tilde{e}_3}$	[0.1,4] TeV	3rd Generation R.h. Slepton Mass
$m_{\tilde{Q}_1}$	[0.4,4] TeV	1st Generation L.h. Squark Mass
$m_{\tilde{Q}_3}$	[0.2,4] TeV	3rd Generation L.h. Squark Mass
$m_{\tilde{u}_1}$	[0.4,4] TeV	1st Generation R.h. up-type Squark Mass
$m_{\tilde{u}_3}$	[0.2,4] TeV	3rd Generation R.h. up-type Squark Mass
$m_{\tilde{d}_1}$	[0.4,4] TeV	1st Generation R.h. down-type Squark Mass
$m_{\tilde{d}_3}$	[0.2,4] TeV	3rd Generation R.h. down-type Squark Mass
$\tan \beta$	[1,60]	Ratio of Higgs VEVs

2.4 Approach

Unlike other dimensional reduction methods, auto-encoders include a mapping back to the original space by design. We utilize a modified Sliced Wasserstein Autoencoder (SWAE) [92] which constrains the shape and consistency of the latent space to be suitable for efficient sampling, unlike standard non-generative auto-encoders. The latent space created is abstract in the sense that there is no direct physical interpretation of its coordinates. But it is also structured such that points are encouraged towards a target distribution using a Sliced Wasserstein loss term. The network finds the most accurate way to encode this structure into reduced dimensionality by minimizing the reconstruction loss. New points that are generated in the low-dimensional latent space are then mapped back up to the original space, allowing for efficient sampling of the valid substructures in high dimensions. All feature columns are normalized to values between 0 and 1 before training.

2.4.1 Why not a VAE?

The Variational Auto-Encoder (VAE) [85] is a well-known generative framework using dimensional reduction, but complications in its structure make it non-ideal for our needs. Notably, the standard analytical form of the VAE’s Kullback-Leibler (KL) loss assumes a Gaussian target distribution, which is achieved when the term is minimized at $\mu = 0$, $\sigma = 1$. This pushes all points in the latent space to within 1 unit of the origin, which when combined with the variational nature of the VAE has been shown to cause information collapse during training: loss values can diverge or become degenerate because KL loss is not a true distance metric [125, 78, 131]. Additionally, because the KL loss term is defined only in terms of μ and σ , it is difficult to fit points to more complicated distributions such as those that are concentric or hollow.

SWAEs come from a family of autoencoders specifically developed to avoid these issues [92, 78], by replacing the KL loss term with the Sliced Wasserstein distance from optimal

transport theory. The Sliced Wasserstein loss term fits latent points to individual points from the target distribution deterministically instead of variationally, mapping directly to latent coordinates instead of μ and σ . Thus SWAEs avoid the VAE’s information collapse problem and can utilize complex target distribution shapes, making them a superior choice for our purposes. Wasserstein-style autoencoders have been shown in the past to produce higher quality samples in image generation tasks, compared to VAEs [92].

2.4.2 SAFESPAM

Using pytorch [113] and an ADAM optimizer [84]², we train a modified SWAE to learn the mappings to and from the latent space, which we denote as f and g respectively. The mapping $\{f : T \rightarrow L\}$ from the theory space T to the latent space L is learned by the encoder, and the return mapping $\{g : L \rightarrow T\}$ is learned by the decoder. The full mapping of the SWAE can then be written:

$$T \xrightarrow{f} L \xrightarrow{g} T. \quad (2.1)$$

In SAFESPAM, the primary modification is the addition of a clustering loss term which encourages valid points to cluster together. The loss function we use thus has three terms: the standard L2 reconstruction loss used by autoencoders to ensure accurate mapping, the Sliced Wasserstein loss term which fits latent points to a target distribution, and the clustering loss term, which pushes valid points toward the origin and invalid points away. Because the clustering loss term is aware of our experimental constraints, SAFESPAM is a supervised deep learning method, although weak-scale values do not enter the model as inputs. Our loss equation takes the following form:

$$\mathcal{L} = \alpha \mathcal{L}_{L2} + \beta \mathcal{L}_{\text{Wasserstein}} + \gamma \mathcal{L}_{\text{Cluster}}. \quad (2.2)$$

²As well as code from the following packages: [86, 61, 127, 74, 102, 128, 80]

where α, β, γ are hyperparameters. Below we explain each term in more depth.

L2 loss, or MSE Mean loss, is calculated between the training data and the output of the SWAE, ensuring that the decoder accurately reconstructs the original inputs when mapping.

It looks as follows:

$$\mathcal{L}_{L2} = \frac{1}{N} \sum_{i=1}^N (t_i - g(f(t_i)))^2, \quad (2.3)$$

where N is the number of points per batch and t the input theory point vector.

The Sliced Wasserstein loss is a formulation of the Wasserstein/earth-mover distance calculated along projected one-dimensional slices in latent space. It quantifies how the target distribution and latent distribution differ from one another. [92, 78, 61]

$$\mathcal{L}_{\text{Wasserstein}} = \frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M tc(\theta_i \cdot \tilde{s}_{k[j]}, \theta_i \cdot f(t_{\ell[j]})). \quad (2.4)$$

Where θ_i represents the randomly sampled one-dimensional slices along which the marginal distributions are defined, $\tilde{s}_k$ and $f(t_{\ell})$ are random samples from the target distribution and the latent input data respectively, and $tc(\cdot, \cdot)$ represents the transport cost between the two distributions.

Our exploratory experiments indicated that for SAFESPAM, strong performance can often be achieved training only on valid points. In these instances, latent valid cores often take a convenient compact shape, useful for sampling and further structural exploration. Heuristically we find that utilizing concentric valid and invalid target distributions outperforms a bimodal normal target distribution where invalid points cluster around a point far from the latent origin.

Our clustering loss term takes inspiration from the Triplet Loss originally used in the fields

of contrastive and metric learning [119, 120], with the goal of embedding a hierarchy of distances in the latent space, such that the distance between a desirable point and anchor point is minimized, and conversely the distance between an undesirable point and anchor point is maximized. For ease of sampling, we choose the latent origin as our anchor point. The mathematical expression per batch for our clustering is expressed as:

$$\mathcal{L}_{\text{Cluster}} = \frac{1}{N} \sum_{i=1}^N \left(a\delta_{\text{valid}}\|f(t_i)\| + \frac{b(1 - \delta_{\text{valid}})}{\|f(t_i)\|} \right), \quad (2.5)$$

where a, b are hyperparameters, N is number of points per batch, and t the input theory point vector. δ_{valid} is 1 for valid points and 0 for invalid points. This term encourages valid points to clusters near the latent origin and pushes invalids outward. When training with only valid points, this term serves as an extra assurance that the valid portion of the latent space is sufficiently compact. When $\gamma = 0$, a SAFESPAM network reverts to an unmodified SWAE.

Together, the Sliced Wasserstein and Clustering loss terms allow us to force favorable structure onto the latent space, controlling the distribution of valid and invalid points. Unlike standard non-generative auto-encoders, where poorly-defined latent spaces make sampling unfeasible, SAFESPAM enables efficient sampling by forcing desirable points to cluster together in one area; see Fig. 2.2 for a cartoon schematic.

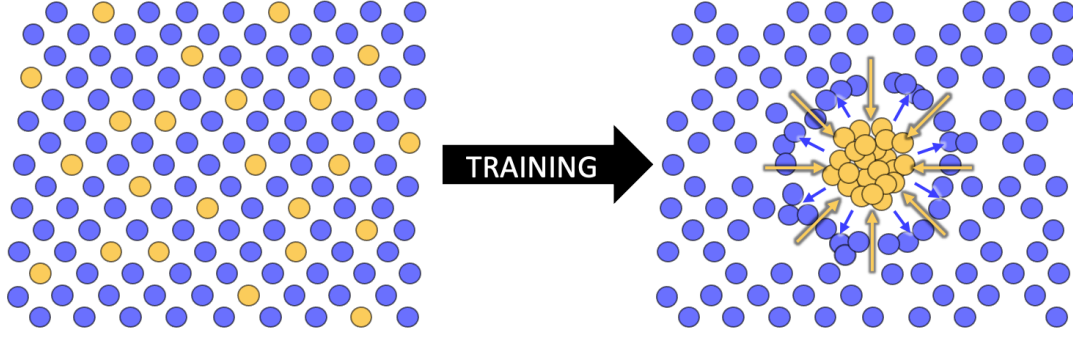


Figure 2.2: A cartoon illustration of the desired outcome of training on the latent space. Minimizing the clustering loss term pushes valid points (shown in orange) towards the latent origin while minimizing the Sliced Wasserstein loss fits those latent valid points to a target distributions. Shown in blue, invalid points (if present in training data) are pushed outward by the clustering loss term. The resulting clustering of valid points in the latent space allows for straightforward, high-efficiency sampling. Minimizing the L2 loss (not pictured) allows the network to successfully map back and forth between this structured latent space and the constrained MSSM parameter space.

2.4.2.1 Sampling the latent space

To sample, we generate points in the latent space and map them back to the original space using the network’s decoder. Our method supports a variety of sampling options, including simple approaches such as uniformly sampling in the valid region. We have found that Kernel Density Estimation (KDE) sampling is best at automatically capturing the shape of the latent valid core. KDE evaluates a kernel function over a dataset to estimate and sample from its probability density function. We utilized a top-hat kernel, with bandwidth h such that $K(x; h) \propto 1$ if $x < h$, to estimate the density of our latent valid cores from known valid points, and sample the latent space accordingly. Using this kernel function, the density at a point in space X in relation to a group of data points x_j can be estimated as

$$\rho_K(X) = \sum_{j=1}^N K(X - x_j, h), \quad (2.6)$$

where N is the number of points being used by the density estimation module. Essentially, this method approximates the density at each location by summing the number of data points

that fall within the bandwidth distance of that point in space. KDE sampling works best in low-dimensional applications, making it well-suited to the latent space of our dimensional-reduction methods.

The valid latent points can be used to build a density estimate, via SciKit Learn’s Density Estimation module [114], from which new points can be sampled. Because our mapping has already done the difficult task of clustering these points together in a reduced-dimensional representation, KDE sampling in our latent space is more efficient than in the original space.

In some cases, the latent valid core takes an irregular shape that is more difficult to sample. To remedy this, we augment our basic KDE sampling approach iteratively. Valid points found in our first round of KDE sampling are added to the pool of points used to fit the next round of KDE sampling; see Fig. 2.3. As this iteration progresses, the bandwidth of the kernel is successively tightened, to sample finer structures more accurately; see Fig. 2.4. Iterative KDE serves as a means of transitioning from initial exploration of the latent space into more intensive sampling of the valid regions already found [44]. This method can be repeated until a satisfactory efficiency is achieved.

While more advanced sampling options exist, such as training a normalizing flow to sample the potentially irregularly-shaped latent valid cores, we find KDE-style sampling to be much less labor-intensive and more robust at handling cases where there are disparities between the location of valid training data and valid sampled data. See, for example, the structures shown in Fig. 2.10.

While SAFESPAM does not guarantee unbiased sampling of the full valid subspace, it does offer significantly higher yields compared with naive sampling. In practice, this often means sampling with very high efficiency over an incomplete portion of the valid range. All SAFESPAM trials shown here used a sampling size of 10k points.

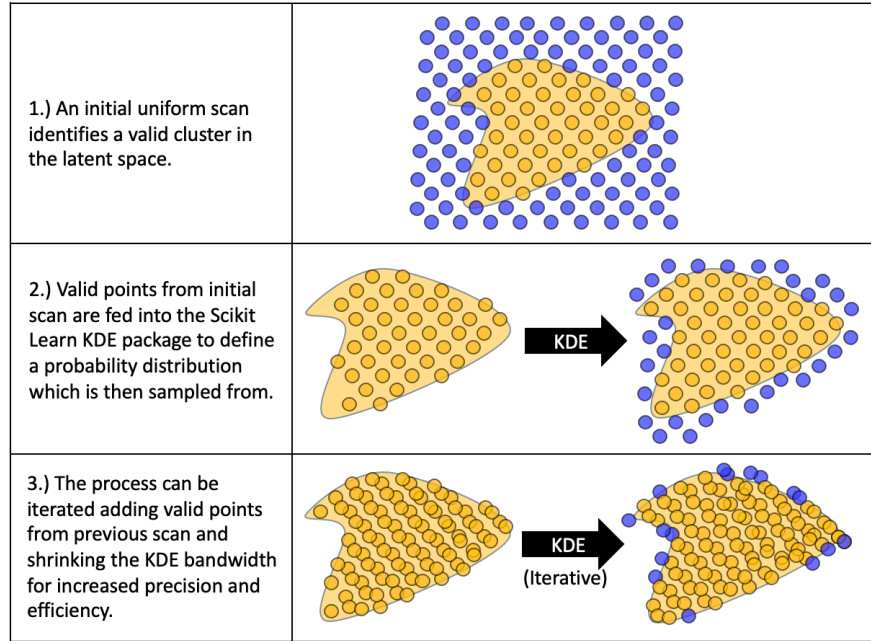


Figure 2.3: A cartoon illustration of the KDE sampling methodology. In CHUNC and CHUNC2, step 1 is omitted and known latent valids from the training data are used instead.

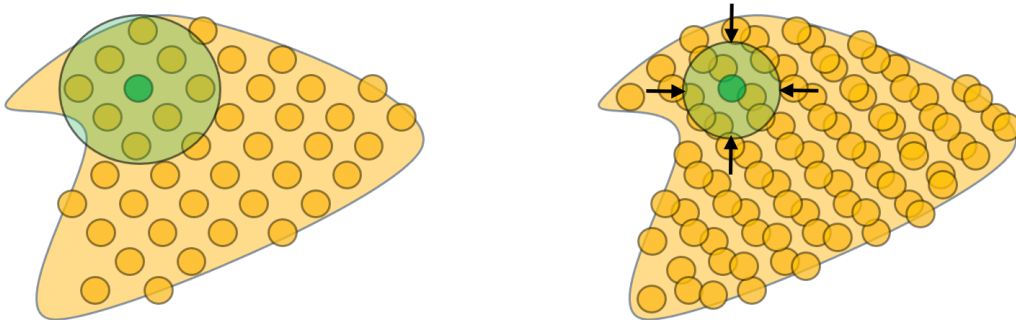


Figure 2.4: A cartoon illustration of the bandwidth shrinking procedure in our iterative KDE sampling methodology. As the number of known valid points grows with each successive iteration, we shrink the bandwidth used to estimate the density at each point in the latent space, allowing for higher precision.

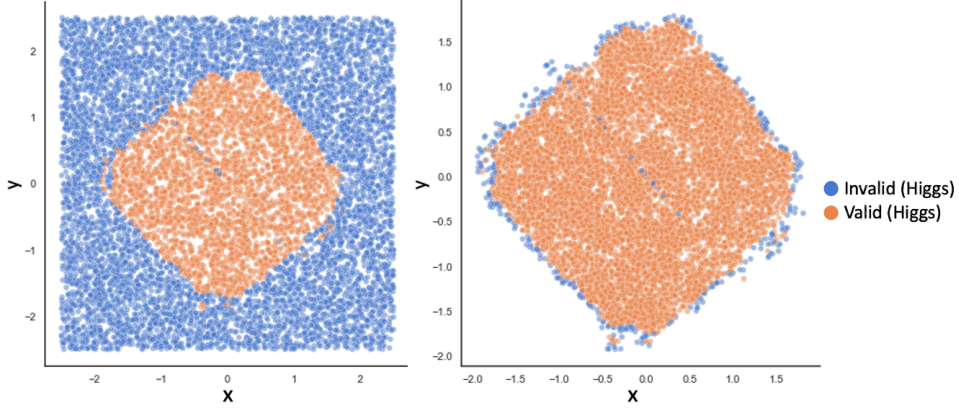


Figure 2.5: An example of KDE sampling in the pMSSM: An initial uniform scan of the latent space (left) demarcates the shape of the valid core, and the valid points found in that scan are utilized via kernel density estimation to sample the core efficiently (right). White spaces represent areas that are avoided by KDE to improve efficiency.

2.4.3 CHUNC

Any function which reduces the dimensionality of its input has the potential to destroy information about the high-dimensional structures it seeks to describe. We anticipate such difficulties in high-dimensional spaces like the CMSSM and PMSSM where collapse to a two-dimensional latent space could cause significant portions of the valid sub-space to be overlooked with SAFESPAM-style methods. Thus, in an attempt to avoid this information loss, we introduce a similar methodology that operates without dimensional compression by imposing that the latent space is of the same dimension as the input space and that none of the layers in the network have a dimension smaller than the input.

The motivation for this approach comes from the Data Processing Inequality (DPI) [43] which states that any function $f : T \rightarrow L$ acting on a space T can only ever destroy or preserve information. This information is represented by the joint probability distribution $p(T, \theta)$, where T is the theory space discussed in Section III and Section IV.2, and θ are the experimental constraints. The SWAE networks constructed here attempt to transform the unknown distribution $p(T, \theta)$ to one whose structure is known $p(L, \theta)$ while preserving

the correlations between L and θ . The DPI states that the mutual information $I[T; \theta] \geq I[f(T); \theta] \geq I[g(f(T)); \theta]$ with equality when no information loss has occurred.

To maximize $I[L; \theta] = I[f(T); \theta]$ then, we impose that the dimension of the space L be the same as the dimension of T . This comes at expense to the ease of sampling which is otherwise gained from utilizing a low dimensional latent space, and we also sacrifice the benefits of visualization. However our latent space still retains the desired quality that valid points cluster near the origin while invalid points are pushed outward. We demonstrate this method in cMSSM and pMSSM on a triple constraint: Higgs, DM, and Neutralino LSP. We train on an equal mixture of valid and invalid points, which we fit to separate non-overlapping target distributions. Our Sliced Wasserstein loss term then becomes:

$$\mathcal{L}_{\text{Wasserstein}} = c\mathcal{L}_{\text{Wasserstein}}^{\text{Valid}} + d\mathcal{L}_{\text{Wasserstein}}^{\text{Invalid}}, \quad (2.7)$$

where c, d are hyperparameters. To reiterate, the role of the latent space is not to generate a low dimensional representation, but rather solely to transform the unknown distribution of the input space $p(T, \theta)$ to one $p(L, \theta)$ which is intentionally structured for sampling purposes, without any potential loss of information. We sample from this known distribution with KDE and map back to the original space, in the same manner as SAFESPAM. Because CHUNC retains usage of the Sliced Wasserstein loss term, the methods avoids the same VAE pitfalls as SAFESPAM. While the lack of dimensional reduction makes CHUNC superficially similar to normalizing flows, we believe our concentric contrastive target distributions are too complex to be utilized in a standard normalizing flow framework, which requires simple and well-understood base distributions [77].

2.4.4 CHUNC2

CHUNC2 is an alternative formulation of CHUNC which posits that further information loss might be avoided if the contrastive clustering is limited to a single additional categorical

variable, θ , allowing the rest of the latent space more freedom to preserve reconstruction capabilities. Instead of pushing invalid points to infinity, the network learns to map to the joint space $L \times \theta$ so that the latent space is constrained as a simple Gaussian $p(L, \theta) = p(\theta)p(L|\theta) = p(\theta)p(\mu_L, \sigma_L|\theta)$ where the categorical variable θ interpolates between signal-like and background-like events. As such, CHUNC2 replaces the clustering loss term mentioned in section 4.2 with a loss term built around an added categorical variable tasked with codifying whether a point is valid or invalid. Thus our latent space is now $N + 1$ dimensional, where the new dimension serves as a predictor of a point’s validity:

- **Binary L^2 Loss** - The latent categorical variable z_i , is constrained through an L^2 loss to match the category $\theta_i = \{s, b\}$ of each event,

$$\mathcal{L}_{\text{bin}}(z, \theta) = \frac{1}{N} \sum_{i=1}^N (z_i - \theta_i)^2. \quad (2.8)$$

In the following studies, the categorical variable is trained to map to 1.0 for valid points and 0.0 for invalid points. It is of course straightforward to generalize this approach to one which θ represents a multi-class problem, but we do not pursue this here since our task is a binary one (i.e. with our triple-constraint task we are not interested in generating points that only meet some but not all of the individual constraints). Sampling utilizes KDE on latent points which fall into the most valid categorical bin to generate new latent points in 6D or 20D respectively, which are mapped back to the original space. Since CHUNC2’s latent space is of a higher dimensionality than the parameter space, normalizing flows would not be able to emulate this methodology.

2.5 Results

2.5.1 SAFESPAM

We trained a series of Modified SWAEs on cMSSM and pMSSM data using Higgs and DM constraints separately. Unless otherwise noted, all SAFESPAM models were trained on valid-only data³, and latent points were constrained to a truncated normal target distribution by minimizing the Sliced Wasserstein loss. The latent space was sampled with an initial uniform scan whose location was determined by the latent position of valid training data. Valid points from this scan were then used to define a probability density sampled with single-round KDE sampling, using a tophat kernel for increased precision. Iterative KDE was utilized to improve results in cases with the Dark Matter experimental constraint. Below, we present the results for Higgs mass constraints in the cMSSM and pMSSM cases, followed by dark matter relic density constraints in each space, and an analysis of the ability of our method to simultaneously satisfy both constraints. We present visuals of the sampled points in the latent space to confirm the presence of a central valid core. We also include histograms comparing our generated valid data to our valid training data to qualify to what extent our sampling has introduced unwanted bias or excluded viable regions of the space. We then report sampling efficiencies, defined as the fraction of points from the sample which adhere to the desired constraint. Relevant hyperparameters are listed in Appendix A. All SAFESPAM model-training for this project took place on a 2017 macbook pro without use of GPU.

We introduce two additional metrics to better quantify the successes and shortcomings of this method, described as follows:

³Note that using valid-only data is not a requirement of the method.

- Core-forming Metric

- This metric measures how successfully the generated valid points match the core shape suggested by the position of latent valid training data. We calculate the Sliced Wasserstein distance between randomly selected points from the latent distribution of valid training points and the latent distribution of valid generated points. Latent points are normalized before calculation to account for differing target distribution sizes between trials. A low value suggests that the generated valid points are occurring where we expect them to be, whereas a higher value suggests potential disparities between the latent distributions of training valids and generated valids. Multiple calculations of this value are averaged out to account for the random nature of the slices. All SAFESPAM metrics were calculated by taking the average over 10 iterations, using 200 projections. It takes the form:

$$\mathcal{L}_{\text{Wasserstein}} = \frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M tc(\theta_i \cdot \tilde{g}l_{k[j]}, \theta_i \cdot f(tl_{\ell[j]})), \quad (2.9)$$

Where θ_i represents the randomly sampled one-dimensional slices along which the marginal distributions are defined, $\tilde{g}l_k$ are the normalized valid latent points generated by the model and $f(tl_{\ell})$ are random normalized latent representations of samples from the valid training dataset, and $tc(\cdot, \cdot)$ represents the transport cost between the two distributions.

- Incompleteness Metric

- This metric measures how successfully the generated valid points account for the full distribution of expected valid points seen in the training data. We calculate the Sliced Wasserstein distance between randomly selected points from the GUT distribution of valid training points and the GUT distribution of valid generated points (excluding valid generated points outside the original range the training

data was generated with). All points are normalized before calculation to account for the different ranges used in data generation in the cMSSM and pMSSM. A low value indicates we are sampling without much bias, whereas a higher value suggests we are introducing more significant bias when sampling. Multiple calculations of this value are averaged out to account for the random nature of the slices. All SAFESPAM metrics were calculated by taking the average over 10 iterations, using 200 projections. It takes the form:

$$\mathcal{L}_{\text{Wasserstein}} = \frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M tc(\theta_i \cdot \tilde{g}g_{k[j]}, \theta_i \cdot f(tg_{\ell[j]})), \quad (2.10)$$

Where θ_i represents the randomly sampled one-dimensional slices along which the marginal distributions are defined, $\tilde{g}g_k$ are the normalized valid GUT points generated by the model and $f(tg_{\ell})$ are random normalized samples from the valid training dataset, and $tc(\cdot, \cdot)$ represents the transport cost between the two distributions.

2.5.1.1 Higgs cases

Our method had little trouble creating a valid core using Higgs mass as the experimental constraint in the cMSSM; see Fig. 2.6, which allowed for highly efficient sampling efficiency (0.957) compared to a naive scan (0.347). Small incursions of invalid points in the core are largely avoided by KDE sampling. The distribution of SAFESPAM-sampled points in the space of cMSSM parameters is shown in Fig. 2.7, where the newly generated points reasonably match the distribution in the original training set.

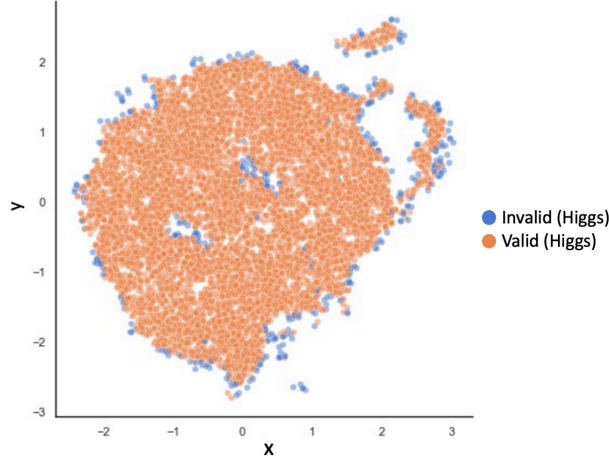


Figure 2.6: Distribution of sampled points in the two-dimensional latent space, for the case of learning to generate cMSSM points which satisfy the Higgs mass constraint. Points which satisfy (fail) the Higgs mass constraint are marked Valid (Invalid). Invalid regions in the center are partially avoided due to KDE sampling.

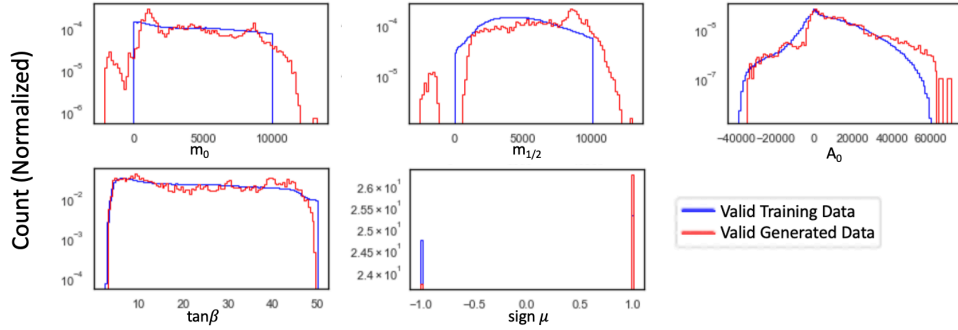


Figure 2.7: Distribution of SAFESPAM-generated points (red) in the cMSSM in each of the five cMSSM parameters, under the Higgs mass constraint, compared to the distribution of the training dataset (blue).

In the case of the pMSSM, SAFESPAM was again able to structure the latent space for efficient sampling. We note that the specific form of the latent space demonstrates significant sensitivity to the hyperparameters such as the shape of the target distribution, length of training, and inclusion or exclusion of valid points, though without a significant change in efficiency.

As a demonstration, we present results for two separate networks trained for the pMSSM

Higgs constraint, labeled Trials 1 and 2 respectively. Trial 1 was trained first on a combination of valid and invalid points fit to a concentric distribution (a normal distribution with an ablated annulus to create two distinct regions), before being trained further with only valid points fit to a truncated normal. Trial 2 was trained only on valid points fit to a truncated normal, but trained for four times as many epochs. Figure 2.8 shows the distribution of the valid points in the latent space for each trial, with clear variation in the structure of the latent space. KDE sampling of the core achieves an efficiency of 0.965 (0.961) for Trial 1 (Trial 2), compared to naive sampling efficiency of 0.171. Distribution of the generated points in the cMSSM parameter space for both trials is shown in Fig. 2.9. Although neither trial generated the full range of valid points seen in the training data, future methods might make use of ensemble sampling to remedy this.

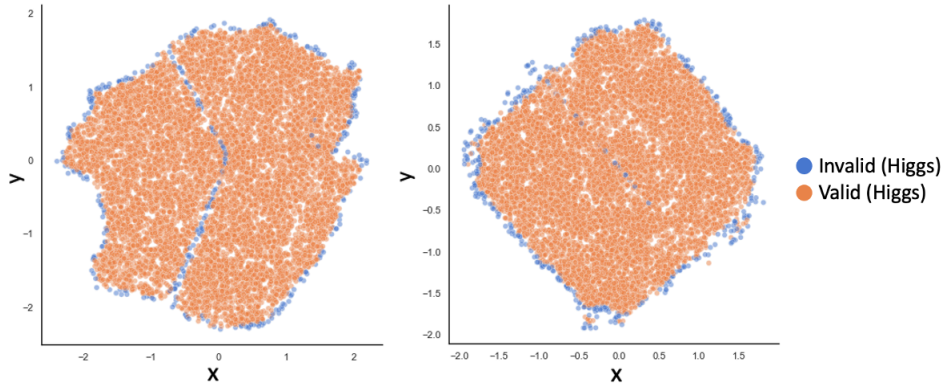


Figure 2.8: Distribution of sampled points in the two-dimensional latent space, for the case of learning to generate pMSSM points which satisfy the Higgs mass constraint. Points which satisfy (fail) the Higgs mass constraint are marked Valid (Invalid). Shown are points sampled via KDE, where invalid regions in the center are partially avoided due to KDE sampling. Trial 1 is shown on the left and Trial 2 on the right, showcasing the sensitivity of the latent space structure to the hyperparameters.

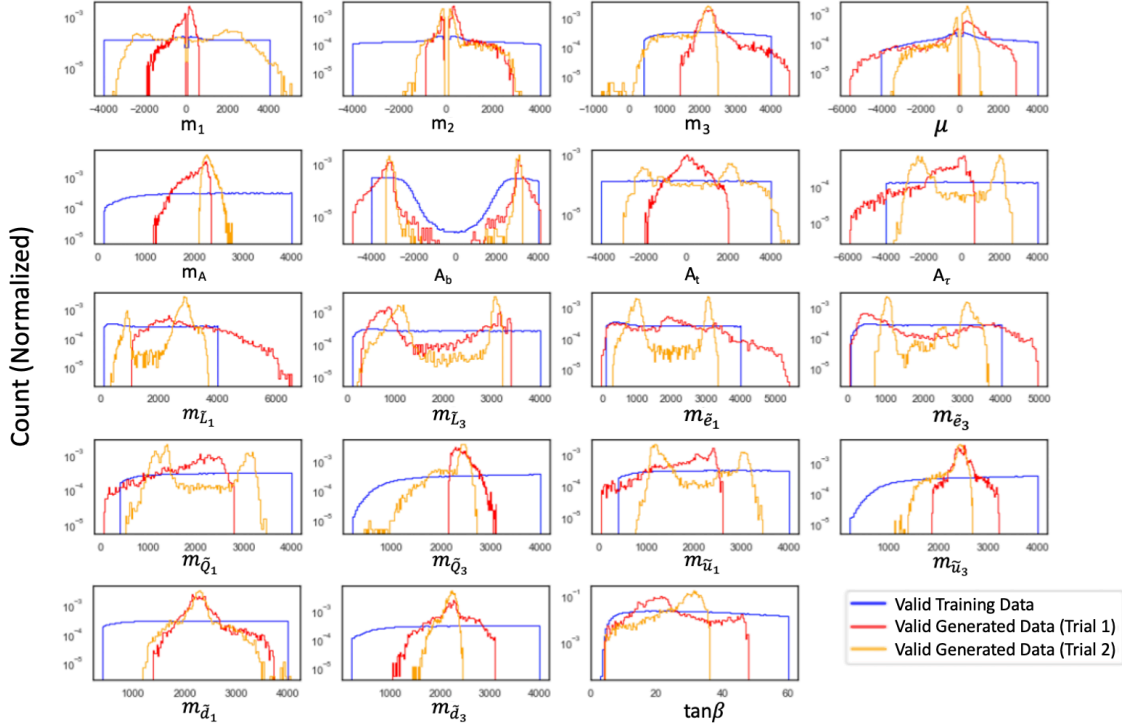


Figure 2.9: Distribution of SAFESPAM-generated points (red for Trial 1, orange for Trial 2) in the pMSSM in each of the nineteen pMSSM parameters, compared to the distribution of the training dataset (blue). While both trials achieve strong efficiency, they sample different portions of the parameter space.

2.5.1.2 Dark Matter Cases

Our dark matter trials had significantly more difficulty creating well-structured latent spaces. While the method was successful in clustering together valid training data, sampling in this vicinity resulted in many points which do not satisfy the constraints; see Fig. 2.10. However, a subset of points from the initial core allowed for efficient generation of new valid points, as discovered after iterating the KDE sampling procedure four times. This led to high efficiency of 0.972(0.979), compared to naive sampling efficiency of 0.05534(0.006506) in the cMSSM (pMSSM). See Figs 2.11 and 2.13. The iterative nature of our sampling method allowed us to probe these valid cores efficiently despite their irregular shape. As in the Higgs trials, this method had more success in preserving the valid range of points from the training data when used in the cMSSM (Fig. 2.12), and less so in the pMSSM (Fig. 2.14). We speculate

that the difficulty here may result from the particularly complex and narrow structures in the cMSSM and pMSSM created by the extreme narrowness of the dark matter constraint as applied to these spaces: we believe our valid-only training dataset lacked the sufficient resolution required to encode their boundaries accurately in the latent space.

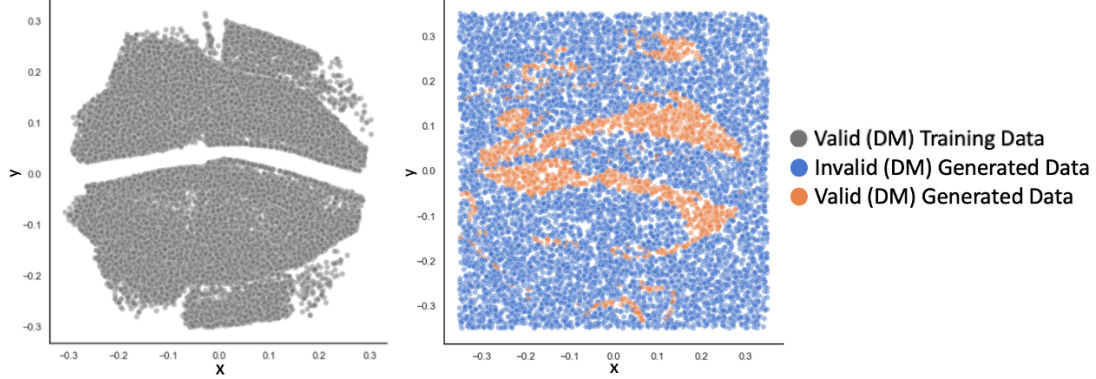


Figure 2.10: Distributions in the two-dimensional latent space for the cMSSM dark matter study, where points which satisfy (fail) the dark matter constraint are marked Valid (Invalid). Left shows the distribution of valid points in the training data, with a well-formed core. However, many new points sampled from this core do not satisfy the dark matter constraint (right), perhaps due to insufficient resolution of DM-valid features in the training data. Successive KDE iterations allowed for precise, efficient sampling in the remaining regions of high validity (see Figs 2.11 and 2.13).

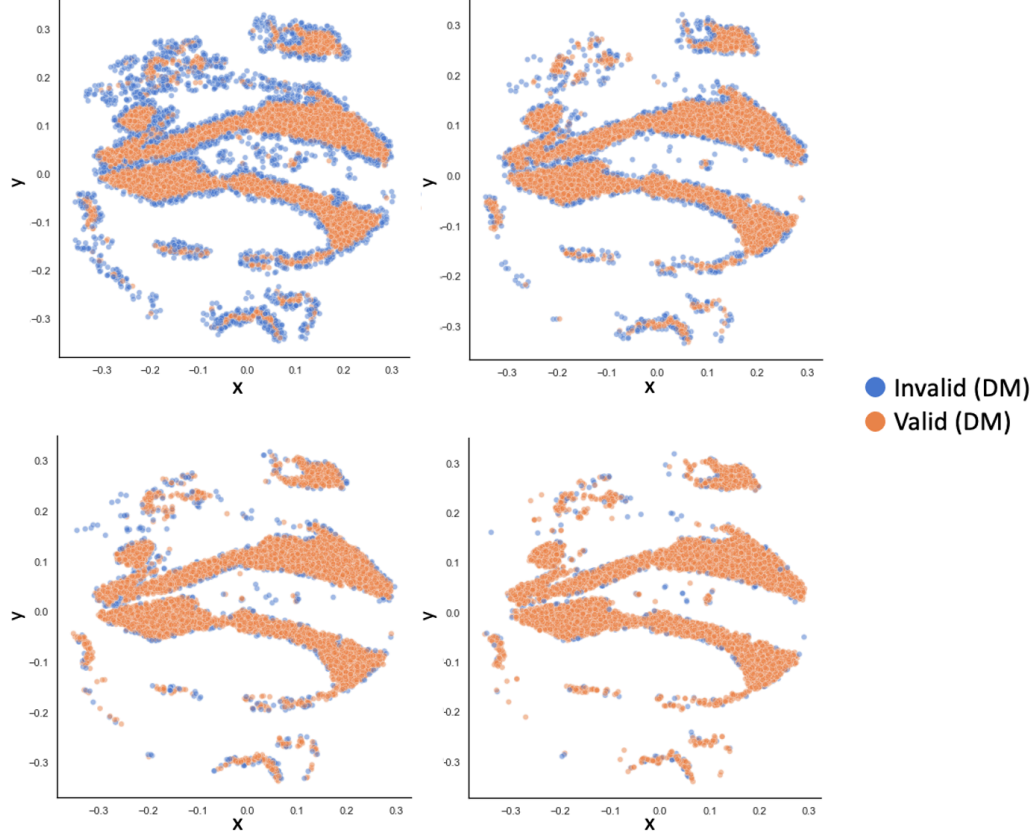


Figure 2.11: Distributions of sampled points in the two-dimensional latent space in four successive rounds of sampling for the cMSSM dark matter case. Points which satisfy (fail) the dark matter constraint are marked Valid (Invalid). In each round, newly discovered valid points are used to update the kernel density estimation of valid regions of the latent space, which is used to perform the next round of sampling. As a result, the ‘crust’ of invalid points shown in blue shrinks away with each successive iteration. The efficiencies for these 4 rounds are 0.724, 0.887, 0.945, and 0.972 respectively.

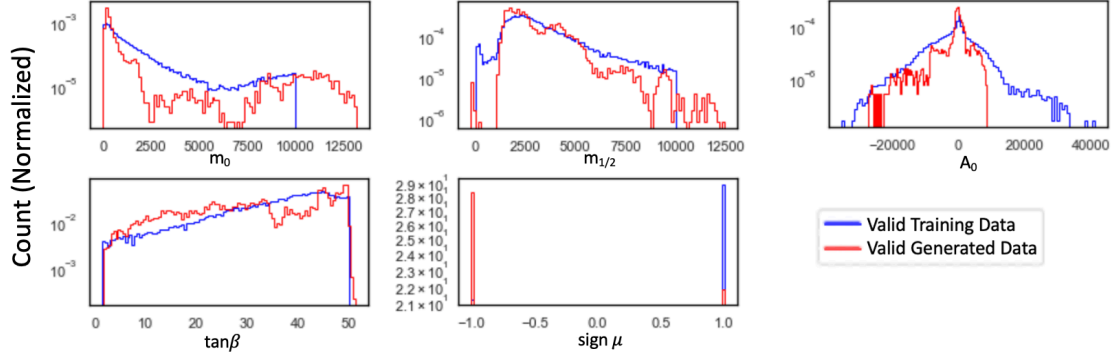


Figure 2.12: Distribution of SAFESPAM-generated points in the cMSSM (red) in each of the five cMSSM parameters under the dark matter constraint, compared to the distribution of the training dataset (blue).

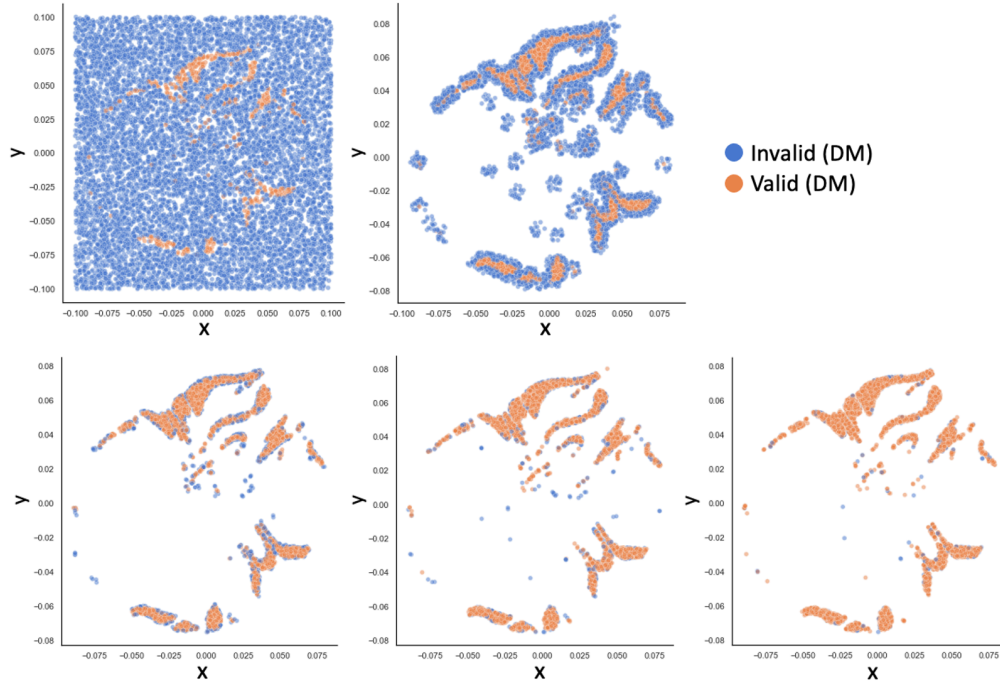


Figure 2.13: Distributions in the two-dimensional latent space for the pMSSM dark matter study, where points which satisfy (fail) the dark matter constraint are marked Valid (Invalid). Shown are distributions of sampled points in successive rounds, where in each round valid points from previous rounds are used to perform a kernel density estimation of valid regions of the latent space for the next round of sampling. The initial scan in the top left frame samples via uniform distribution over a region determined by the position of latent valid training data. The efficiency for each round is 0.049, 0.432, 0.881, 0.950, 0.979, respectively.

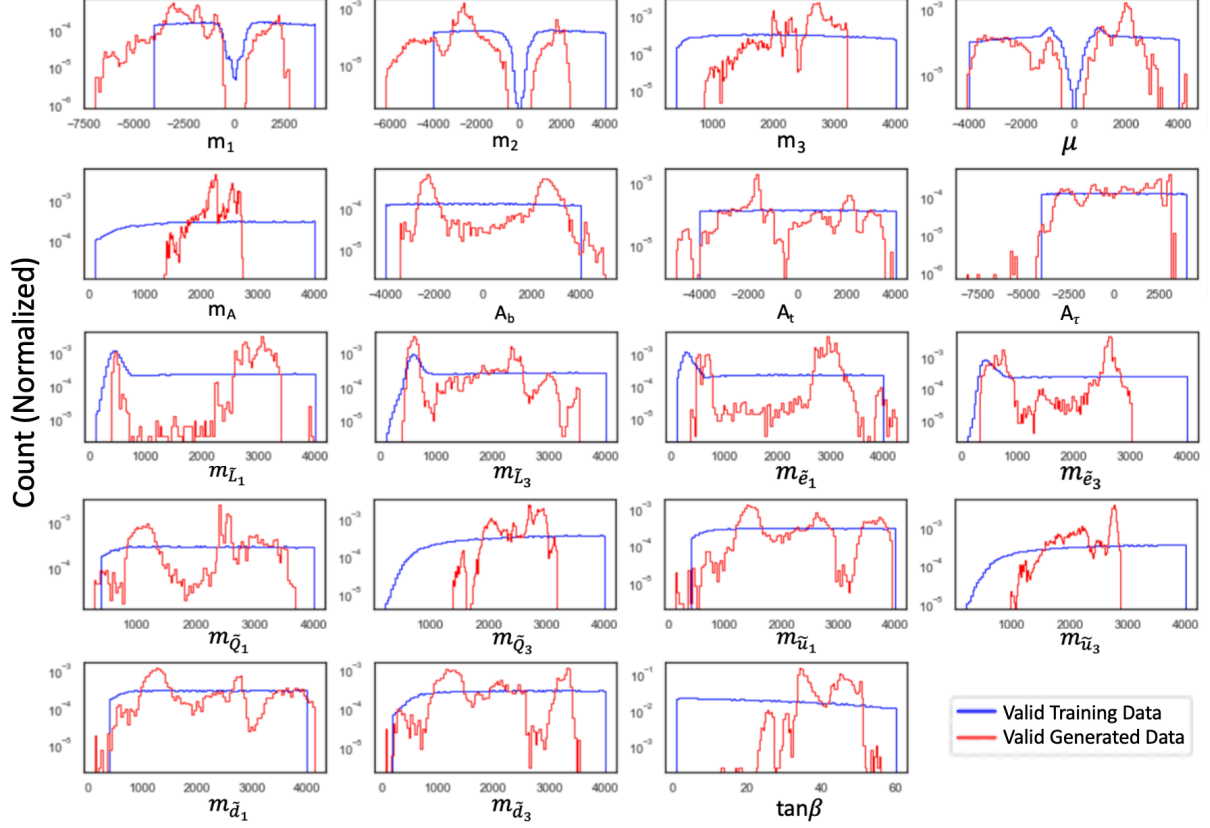


Figure 2.14: Distribution of SAFESPAM-generated points in the pMSSM (red) in each of the nineteen pMSSM parameters under the dark matter constraint, compared to the distribution of the training dataset (blue).

2.5.1.3 Adding secondary constraints without additional training

The structure of the SAFESPAM latent space allows us to apply additional constraints *in media res* when sampling. We identify latent points which meet these additional constraints via initial scan or previous sampling and use KDE sampling to draw new points from their vicinity in the latent space, without providing any additional information to the auto-encoder about which points satisfy these secondary constraints. We demonstrate this in the cMSSM (Fig. 2.15) and pMSSM (Fig. 2.16) by re-sampling our Higgs models for points that have both valid Higgs and DM values, i.e. a joint constraint. Although our method does not guarantee that such ‘double-valid’ points will exist in the latent space or form any neat self-contained substructure, we found our KDE sampling was sufficiently robust to sample the

clusters of points that were present in the latent space, achieving a significantly increased efficiency compared to the naive scan, and doing so without requiring any additional training. This property gives our maps an added versatility that many non-mapping generative models and efficient search methods are unequipped for. For example, one can add additional experimental constraints or look for theory points from a specific dark matter annihilation class just by selectively sampling regions of the map where those characteristics are known to exist. Our method’s dimensional reduction capabilities also give it a significant advantage in ease of sampling these added constraints compared to other mapping-based generative methods such as normalizing flows which could in theory attempt something similar. We see significant increases in efficiency with single-round KDE up to 0.3191 (0.5621) in the cMSSM (pMSSM) compared to a naive value of 0.004467 (0.01892). Additional iterations of KDE may improve these results further.

Table 2.3 summarizes the generation efficiency of experimentally valid points using SAFES-PAM and naive scanning.

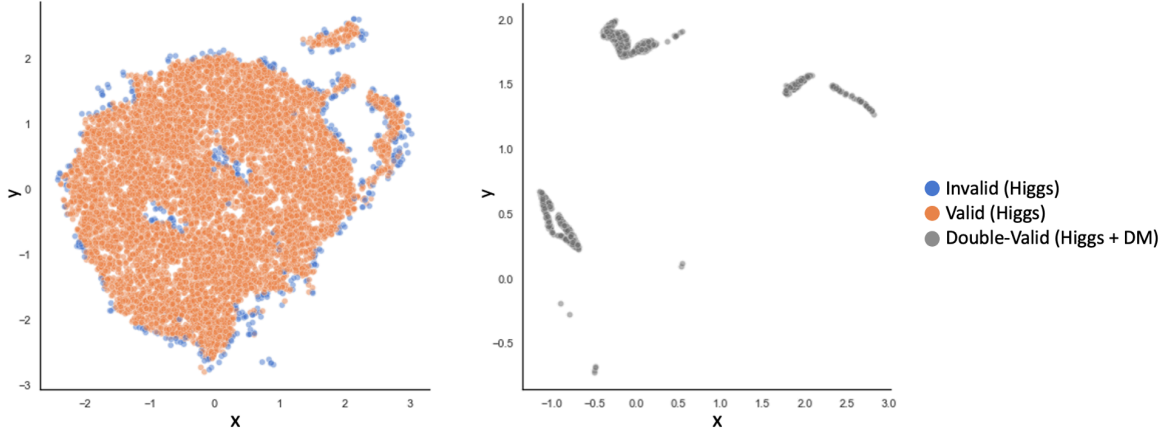


Figure 2.15: Distributions in the two-dimensional latent space for the cMSSM study, where points which satisfy (fail) the Higgs mass constraint are marked Valid (Invalid), and valid points which also satisfy the dark matter constraint are marked as Double-Valid. In the left, the initial scan identifies a core of valid Higgs points. On the right are the newly generated points which satisfy both constraints.

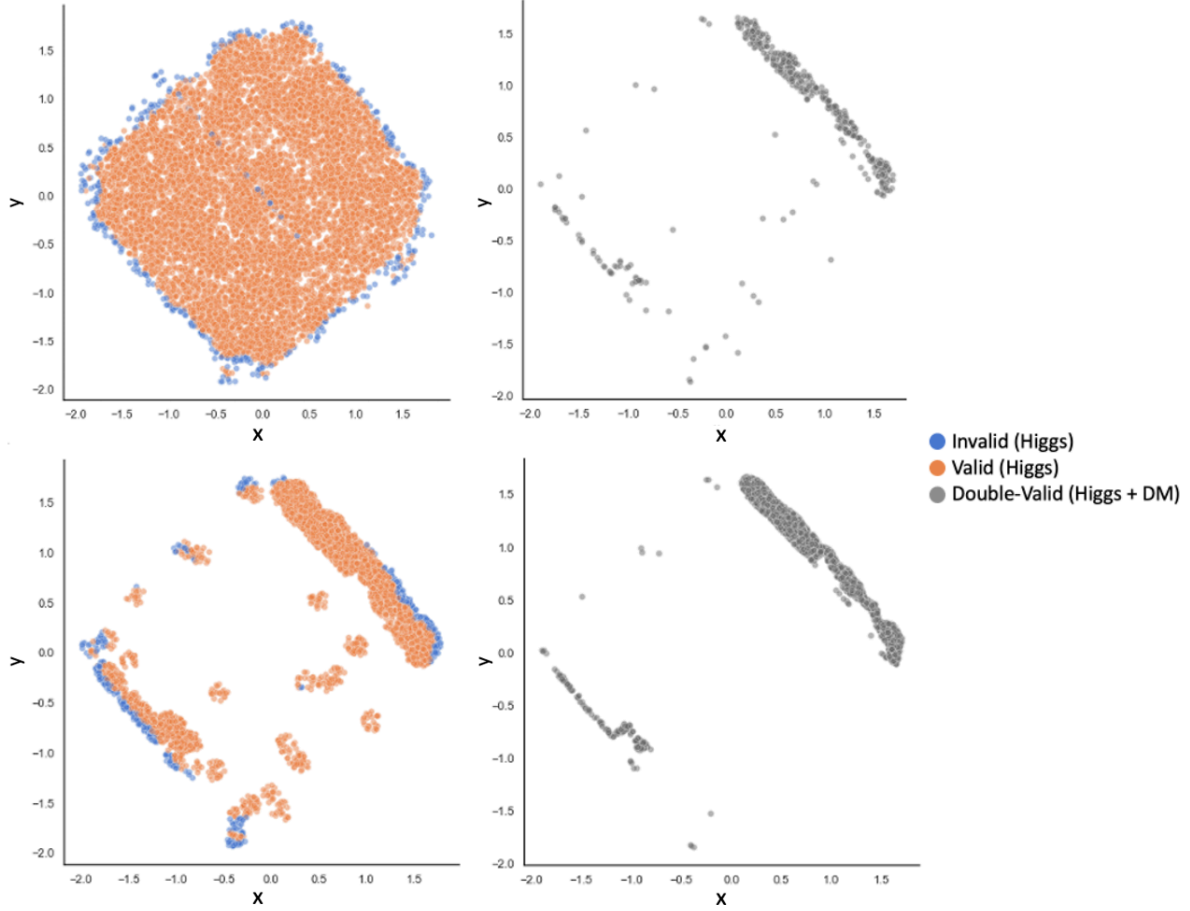


Figure 2.16: Distributions in the two-dimensional latent space for the pMSSM study, where points which satisfy (fail) the Higgs mass constraint are marked Valid (Invalid), and valid points which also satisfy the dark matter constraint are marked as Double-Valid. In the top left, the initial scan identifies a core of valid Higgs points. In the top right, the subset of those points which also satisfy the dark matter constraints are identified. In the bottom left are points generated using a KDE built from the double-valid points, and in the bottom right are the new points which satisfy both constraints.

Table 2.3: Comparison of the efficiency of generation of experimentally-valid points in naive scanning versus SAFESPAM, for several experimental constraints (Higgs mass m_H , dark matter relic density Ω_{DM} , or both) and two theoretical spaces (5-dim cMSSM, 19-dim pMSSM).

Space	Constraint	Method		
		Naive scan	SAFESPAM	Notes
cMSSM	m_H	0.347	0.957	
pMSSM	m_H	0.171	0.965	Trial 1
			0.961	Trial 2
cMSSM	Ω_{DM}	0.0065	0.972	Iterative
pMSSM	Ω_{DM}	0.0553	0.979	Iterative
cMSSM	$m_H \cap \Omega_{DM}$	0.0045	0.319	Trained for m_H
pMSSM	$m_H \cap \Omega_{DM}$	0.0189	0.562	Trained for m_H

Table 2.4: Comparison of the Core-Forming and Incompleteness Metrics for various SAFESPAM trials for several experimental constraints (Higgs mass m_H , dark matter relic density Ω_{DM}) and two theoretical spaces (5-dim cMSSM, 19-dim pMSSM).

Space	Constraint	Core-Forming	Incompleteness	Notes
cMSSM	m_H	0.0037	0.0043	
pMSSM	m_H	0.0167	0.0202	Trial 1
		0.0081	0.0135	Trial 2
cMSSM	Ω_{DM}	0.0051	0.0152	Iterative
pMSSM	Ω_{DM}	0.0403	0.0355	Iterative

2.5.2 CHUNC and CHUNC2 for Triple Constraint

The cMSSM and pMSSM training datasets for CHUNC and CHUNC2 each consist of approximately 10K points, evenly split between invalid points and triple-valid points, which satisfy the Higgs mass, dark matter relic density and LSP constraints defined in section 2.

The higher-dimensional latent space requires a slightly different KDE sampling method. Rather than performing a uniform scan to identify a potential core, we use the position of known latent valid points from the training set to perform the kernel density estimation. CHUNC attempts to cluster all invalid points on an N -dimensional spherical shell of radius 1 with thickness of 0.6, and valid points to an N -dimensional spherical core of radius 0.3. In CHUNC2, all points are forced to a Gaussian distribution in the original dimensions, while the additional categorical quantity is constrained to a value of 1.0 (0.0) for valid (invalid) points.

Despite not using dimensional reduction, CHUNC and CHUNC2 both achieve more than two orders of magnitude improvements in generation efficiency over the naive scan, see Table 2.5. In the cMSSM, CHUNC2 has an efficiency of 0.32, outperforming CHUNC's efficiency of 0.21. However, in the higher dimensional pMSSM case CHUNC's efficiency of 0.24 exceeds CHUNC2's efficiency of 0.15. It is difficult to say why this is the case. Iterative KDE sampling may improve these results further.

While CHUNC and CHUNC2 appear evenly matched at preserving the proportionality of the valid subspace in the lower dimensional cMSSM trials (see Fig. 2.17), CHUNC2 does a visibly better job at capturing the proportionality of the valid subspace without introducing significant distortion, as seen most notably in the $|M_2|$, $|A_t|$, $m_{\tilde{L}_1}, m_{\tilde{L}_3}$ and $m_{\tilde{u}_3}$ distributions in Fig. 2.18. However, CHUNC outperforms CHUNC2 for pMSSM efficiency. However our 'incompleteness' metric suggests these differences may be less prominent than they appear. Both CHUNC and CHUNC2 appear to introduce notably less bias than SAFESPAM when sampling from the pMSSM.

Table 2.5: Comparison of the efficiency of generation of experimentally-valid points in naive scanning versus CHUNC and CHUNC2 methods for three experimental constraints (Higgs mass m_H , dark matter relic density Ω_{DM} , and the lightest supersymmetric particle (LSP) being a neutralino) and two theoretical spaces (5-dim cMSSM, 19-dim pMSSM). The uncertainty across all trials is $\sim \mathcal{O}(0.01)$, estimated by running 10 independent samplings for each case.

Space	Constraint	Method		
		Naive scan	CHUNC	CHUNC2
cMSSM	$m_H \cap \Omega_{DM} \cap \text{LSP}$	0.0004	0.21	0.32
pMSSM	$m_H \cap \Omega_{DM} \cap \text{LSP}$	0.0012	0.24	0.15

Table 2.6: Comparison of the Core-Forming and incompleteness Metrics for CHUNC and CHUNC2 for the 'triple' experimental constraint (Higgs mass $m_H \cap$ dark matter relic density Ω_{DM} *cap* Neutralino LSP,) in two theoretical spaces (5-dim cMSSM, 19-dim pMSSM). Note that CHUNC2's categorical variable is excluded from calculations of the Core-Forming metric.

Space	Method	Constraint	Core-Forming	Incompleteness
cMSSM	CHUNC	$m_H \cap \Omega_{DM} \cap \text{LSP}$	7.9545	0.0252
pMSSM	CHUNC	$m_H \cap \Omega_{DM} \cap \text{LSP}$	0.0212	0.0023
cMSSM	CHUNC2	$m_H \cap \Omega_{DM} \cap \text{LSP}$	0.0163	0.0174
pMSSM	CHUNC2	$m_H \cap \Omega_{DM} \cap \text{LSP}$	0.0007	0.0020

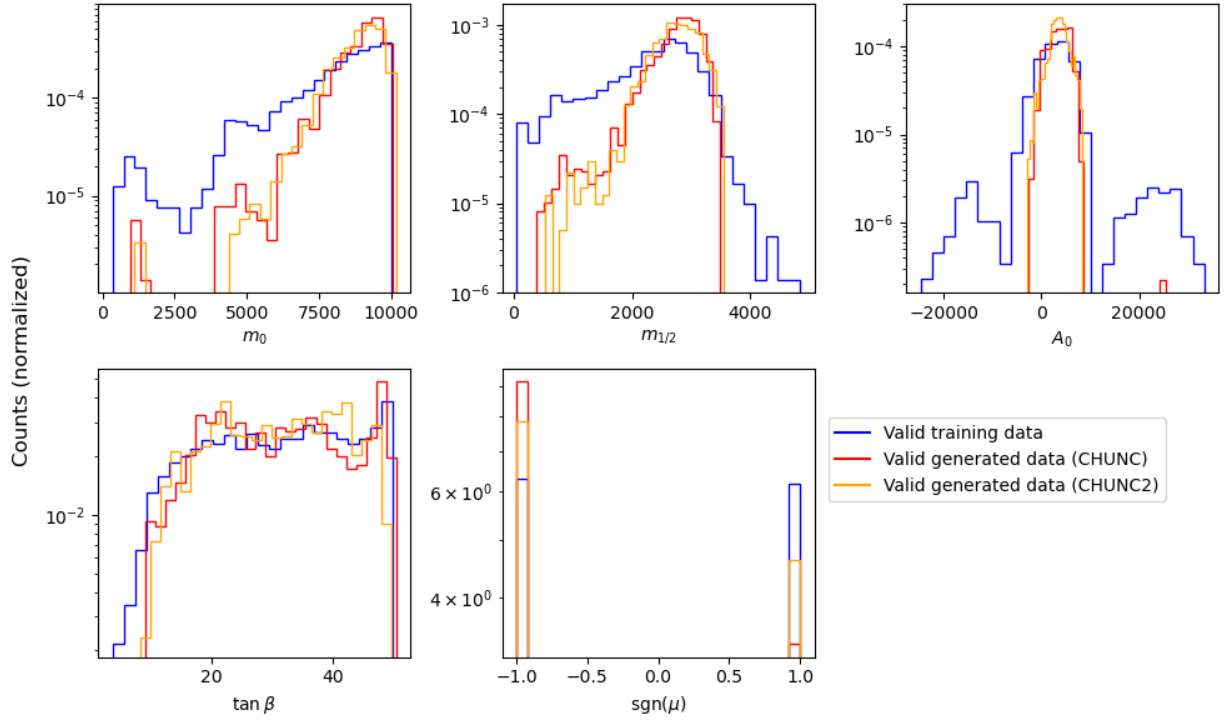


Figure 2.17: Distribution of CHUNC (red) and CHUNC2 generated (orange) points in each of the five cMSSM parameters under the triple constraint, compared to the distribution of the training dataset (blue).

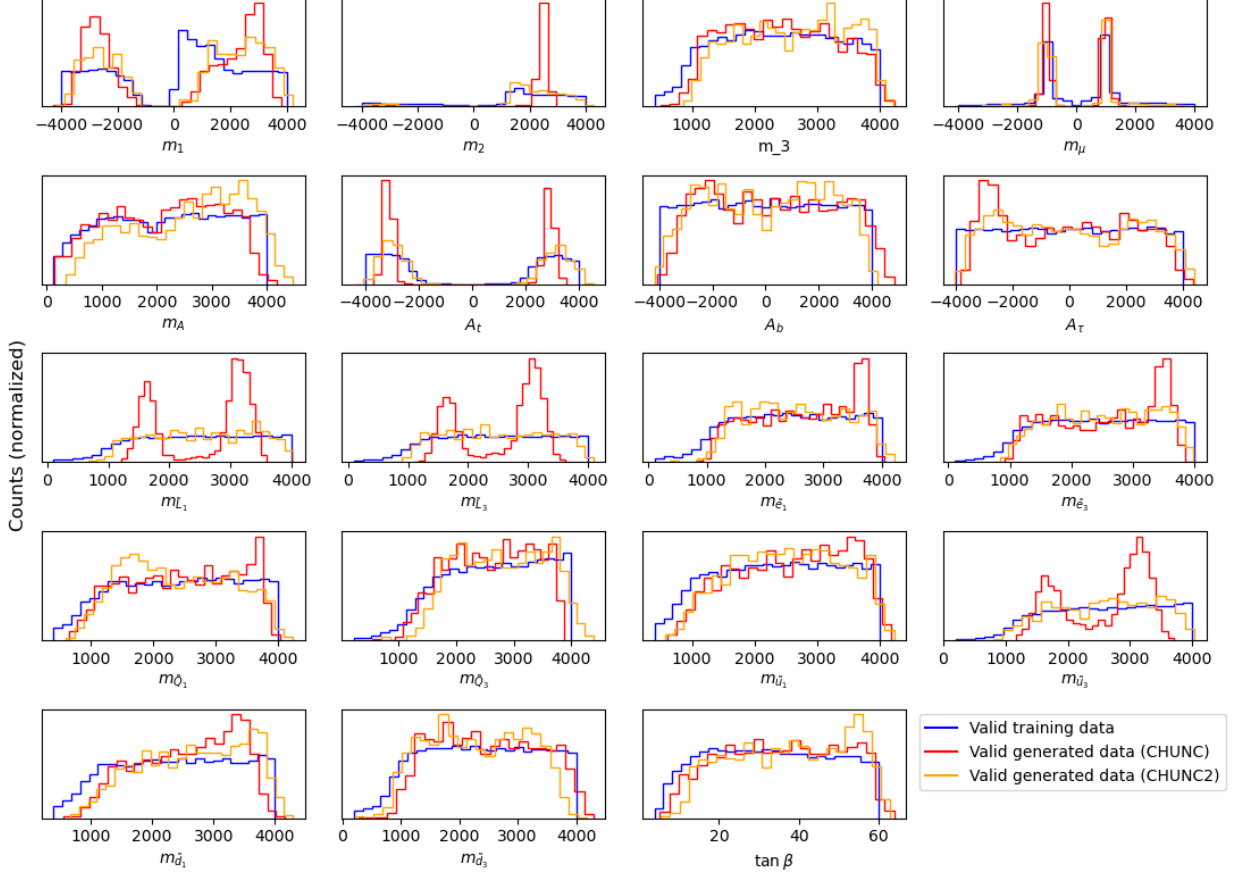


Figure 2.18: Distribution of CHUNC (red) and CHUNC2 generated (orange) points in each of the nineteen pMSSM parameters under the triple constraint, compared to the distribution of the training dataset (blue).

2.6 Discussion and Conclusion

2.6.1 Strengths and Weaknesses

SAFESPAM is capable of significant improvements in efficiency over naive sampling, attaining greater than 95% efficiency in all single-constraint cases; see Table 2.3. However, the range of points sampled often missed valid regions from the training data, especially in the pMSSM. In many cases we also found valid points generated outside the boundaries specified in our initial data generation, in ranges previously unknown to the model, behavior not unexpected for such models. For example, Fig. 2.7 shows valid points with $m_0, m_{1/2} > 10,000$

GeV. While the latent distributions produced by SAFESPAM often display a distinct boundary between valid and invalid points, this was not the case for the Dark Matter trials, perhaps due to the narrower constraint in weak-scale space.

Differences in evaluation software, differing combinations of experimental constraints, variations on which parameters of the spaces are being explored and the range of values over which they are explored all make it a significant challenge to make direct comparisons to other state of the art methods. In most cases these differences constitute essentially different though related problems with potentially significant differences in level of difficulty. That said, if we ignore this we find that for the Higgs mass constraint, in both the cMSSM and pMSSM SAFESPAM achieves efficiencies which significantly outperforms prior work using HMC (0.723 cMSSM and 0.319 pMSSM), normalizing flows (0.796 cMSSM and 0.663 pMSSM), NSGA-II (0.715 cMSSM and 0.862 pMSSM) and TPE (0.668 cMSSM and 0.557 pMSSM) methods, and slightly outperforms CMA-ES (0.924 cMSSM and 0.899 pMSSM) [77, 44]. A significant caveat is that SAFESPAM often fails to capture portions of the valid subspace, especially in the pMSSM. Other methods do not seem to struggle with this to the same extent (although this is difficult to quantify since few papers provide a numerical metric which directly measures this sort of sampling bias in aggregate over all dimensions).

Across SAFESPAM, CHUNC, and CHUNC2 individually, we see that better (lower core-forming) values seem to generally correspond with better (lower incompleteness), but comparing across methods suggests that the removal of the constraints of dimensional reduction allows for greater leeway in this regard: In SAFESPAM core-formation seems to become more difficult as the number of dimensions increases, but for CHUNC and CHUNC2 the opposite holds true. Although CHUNC seems to struggle with core formation at times even more than SAFESPAM, this does not seem to effect its incompleteness nearly as much, perhaps because of the method’s lack of dimensional reduction. Notably in the cMSSM CHUNC has a core-forming value several orders of magnitude worse than any other trial, but is still

able to achieve better incompleteness than our pMSSM SAFESPAM dark matter trial. With respect to core-forming alone, CHUNC2 seems to perform around 2 orders of magnitude better than corresponding CHUNC trials in both the pMSSM and cMSSM respectively, perhaps because the contrastive task has been shifted onto the categorical variable. However it is interesting and worthwhile to note the lack of pronounced effect this core-forming issue has on incompleteness in CHUNC and CHUNC2, compared to SAFESPAM. It is also possible that the differences in incompleteness we observe are due to CHUNC and CHUNC2 being trained on a combination of valid and invalid points where SAFESPAM was trained only on valid data. Additionally we see that although CHUNC2 outperforms CHUNC in both the core-forming and incompleteness metrics in both spaces, this does not always lead to better efficiency. It is difficult to quantify definitively why this is the case.

Both SAFESPAM and CHUNC-style methods show notable improvement across several constraints compared to naive sampling in the cMSSM and pMSSM. The ability to apply additional constraints while sampling makes these maps versatile tools, and the iterative usage of KDE sampling guarantees that even difficult latent shapes can be sampled with high efficiency. Optimistically, one imagines a web of theorists training SWAE networks on specific sub-regions of interest in these parameter spaces and sharing their models with one another for easy data generation in those regions. Another possibility is to use these methods to explore lesser-known and under-explored subspaces within the full MSSM, such as those which might lead to new solutions to the little hierarchy problem [59, 12]. Our methodology could also be used with non-MSSM theoretical parameter spaces, or any high-dimensional supervised learning problem with a clear labeled distinction between valid and invalid points.

Broadly, there is a clear trade-off between the efficiency benefits of dimensional reduction methods such as SAFESPAM and the potential sampling biases introduced by compressing these high-dimensional spaces. In contrast, CHUNC and CHUNC2-style networks avoid dimensional reduction in their latent spaces, and achieve significantly less bias in sampling

the pMSSM despite imposing no constraints that select for this, allowing these methods to sample a broader range of valid points. However, with no dimensional reduction, these latent spaces are still vast and difficult to search exhaustively, even if well-organized, and KDE-style sampling can be less effective, resulting in lower yield.

2.6.2 Future Avenues

In terms of near term improvements of our methodology, sampling jointly from an ensemble of models with varying hyperparameters could help recover portions of these spaces lost due to dimensional reduction in any given individual model. Alternatively, one might introduce a secondary Sliced Wasserstein loss term, applied to fit randomly generated output data to the expected proportionality of the valid sub-space while training, in order to prevent biased sampling.

The cMSSM and pMSSM both adopt de facto theory biases in constraining the underlying MSSM space. Ideally, machine-learning-driven methods will someday be powerful enough to tackle larger unconstrained spaces with high accuracy and efficiency, usurping the need for such constraints. However neither the SAFESPAM or CHUNC-style method in their current iterations would perform well in a ~ 100 parameter space such as the unconstrained MSSM. Reduction to two dimensions via a SAFESPAM-style methodology would cause significant information loss, and a CHUNC-style method with no dimensional reduction would be overwhelmed by the sheer vastness of the space, the primary reason the MSSM is difficult to explore in the first place. Standard KDE sampling is also known to degrade due to curse of dimensionality, though there are proposed methods to circumvent this [106, 45, 115]. Ultimately, future work in this vein will need to thread the needle with dimensional reduction in order to simplify sampling and search methods without incurring significant information loss. One optimistically imagines a variable-dimension auto-encoder where the dimension of the latent space at any given point is itself a trainable parameter. Such a network could

constrain different portions of a space into separate yet interconnected lower-dimensional representations, changing the dimension of the latent space as needed to maximize dimensional reduction benefits while avoiding information loss. The boundaries of the resulting structures in the latent space could be elucidated iteratively then those structures could be learned and sampled via manifold learning [112, 22].

Another fundamental issue is that as dimensionality increases a larger initial scan is needed to represent the space without losing resolution. In the unconstrained MSSM any sort of uniform scan is currently intractable: a scan with only 2 values per dimension would require $\sim 2^{100}$ points. A potential solution to this would be to focus on sequential or algorithmic search methods that can operate without the overhead of evaluating a large initial dataset (see [104], [3] for examples), but it is unclear how exactly one would meld these with a dimensional-reduction-based framework.

2.6.3 Conclusion

We trained a variety of modified Sliced Wasserstein autoencoders, both with and without dimensional reduction, to sample efficiently from experimentally valid regions of two subspaces of the MSSM, the cMSSM and pMSSM, under multiple experimental constraints. Utilizing these mappings alongside KDE sampling, we were able to demonstrate efficiencies significantly greater than naive values in all cases, and we suspect additional iterative usage of KDE sampling could improve these results even further. Our method also is capable of easily adding additional secondary constraints without any retraining, in part due to the benefits of dimensional reduction.

If subtle, unexplored regions remain in spaces like the cMSSM and pMSSM, their discovery hinges on sampling methods that can overcome the inherent difficulties of searching high-dimensional spaces. Methods like SAFESPAM, CHUNC, and CHUNC2 provide a novel way to drill deeper into such parameter spaces and ensure underexplored regions do not

go overlooked. Further work with these methods to analyze such regions in the cMSSM and pMSSM remains, as do additional iterative strategies to apply additional successive experimental constraints. The discovery of notable regions in these spaces could easily inspire to new LHC analyses in search of supersymmetric signatures we would not know to look for otherwise.

Chapter 3

Iterative Flows: Towards an exhaustive search of the PMSSM

Abstract— We utilize normalizing flows to perform an iterative search of the 19-dimensional pMSSM space across multiple experimental constraints. After learning the distribution of theory points which meet a single constraint, we generate a new dataset which we use to train on the distribution which meets 2 simultaneous constraints, and so on. While most searches are forced to choose between going "deep" (many constraints) or "wide" (more coverage of the search space), our iterative methodology allows us to bridge the gap and apply these constraints jointly to conduct a thorough search without loss of resolution.

3.1 Introduction

In the growing field of Machine Learning (ML) for High Energy Physics (HEP) theory applications, one common task is efficiently searching through a large 'parameter space' for individual theories which agree with experiment [32, 96, 23, 77, 104, 4, 44, 3, 112, 22, 97, 98, 14, 13]. This is difficult because the theory parameters are expressed at the Grand

Unified Theory (GUT) scale and the experimental observables are expressed at the weak scale. While there exists a mapping from GUT to weak, experimental observables cannot be simply mapped onto GUT parameter spaces. This difficulty is compounded by the high-dimensionality of many HEP parameter spaces. For instance the Minimal Super-Symmetric Model (MSSM) has >100 parameters [100]. Broadly speaking, these searches can currently be divided into two categories:

- ‘Search-as-you-go’ methods which explore theory space *sequentially* without the overhead of a large initial scan, but do so with limited breadth [104, 14, 3, 38, 75, 109, 130, 34, 72, 108, 39, 31]
- ‘Proportional’ methods ¹ which perform high-dimensional interpolation to generate theory points from a desired distribution, but require a typically large initial dataset to accurately do so [77, 112, 22, 97, 98, 44].

While ‘as-you-go’ searches are highly adept at finding viable models in ‘needle-in-a-haystack’ contexts like previously unexplored spaces, they are not designed to exhaustively search through theory-space, and cannot capture the full breadth of desirable points across various sub-regions without incurring significant bias. To do so requires a methodology which can capture the underlying proportionality of the desired subspace to ensure that promising regions are not overlooked: Only proportional methods can approximate an initial distribution this way. If one is satisfied with finding *any* good model, as-you-go methods will suffice, but to find *every* good model—or even rule out a search space’s viability entirely—requires a proportional methodology.

The accuracy of such proportional methods depends heavily on the initial training data, which is a significant limitation: Biased or sparse training data can inadvertently teach an

¹Naively, one might define a method as ‘ideally proportional’ if it generates points which minimize the earth-mover or Wasserstein distance [117, 65] with respect to the desired target distribution. However this is complicated in our case because our methodology seeks explicitly to discover regions of the parameter space which a naive scan has missed due to lack of resolution.

ML model to ignore promising regions of search space. **Crucially**, this poses particular difficulty when attempting to apply multiple constraints in parallel: the more one constricts, the sparser and less reliable data from that initial scan becomes in representing the true nature of the multi-valid subspace. In high dimensional spaces, where initial scans often lack the sufficient resolution to adequately represent such complex distributions, the small, irregular or isolated pockets containing desirable models can easily be overlooked.

However, using such methods *iteratively* presents an appealing solution to this problem: since new datasets can be generated between each additional constraint, the search-resolution is not diluted. Building on past work [77], we perform an iterative search of the Phenomenological MSSM (pMSSM), a 19-dimensional subspace of the MSSM [50, 51].

Because immense spaces like the full MSSM are essentially intractable to search, the pMSSM acts as a simplified, theory-motivated intermediary, by enforcing three high-level constraints: No new sources of CP violation, no flavor-changing neutral currents, and first-second generation universality [50]. These limitations exclude phenomenologically ill-behaved portions of the MSSM while also reducing the over-all dimensionality and allowing searches to take place [50]. While tightening experimental constraints [4, 5] and lack of direct detection at colliders [55] have gradually disfavored the PMSSM amongst other possible theories, it still plays a significant role in guiding current LHC experiments and inspiring new analyses [64, 73, 9, 25, 23, 29, 30, 2, 82, 46] making it very relevant to present-day HEP physics.

We proceed with our iterative search of the pMSSM as follows: Instead of learning multiple experimental constraints in parallel, we do so in series. Once a constraint is learned with the correct proportionality, a new dataset is generated and the cycle repeats. We utilize normalizing flows for this task, a generative method known for its reliability in reproducing the proportionality of high dimensional data [88], to generate points which satisfy a multitude of constraints simultaneously. Importantly, the iterative nature of our search method ensures that the space is probed without loss of search-resolution, minimizing the possibility of

missing small or isolated but promising regions of theory space.

We proceed as follows: In section 2 we describe the task and define our experimental constraints and detail our initial dataset generation procedures. In section 3 we describe the specifics of our iterative approach. In section 4 we present and discuss our results. Section 5 consists of additional discussion and conclusions.

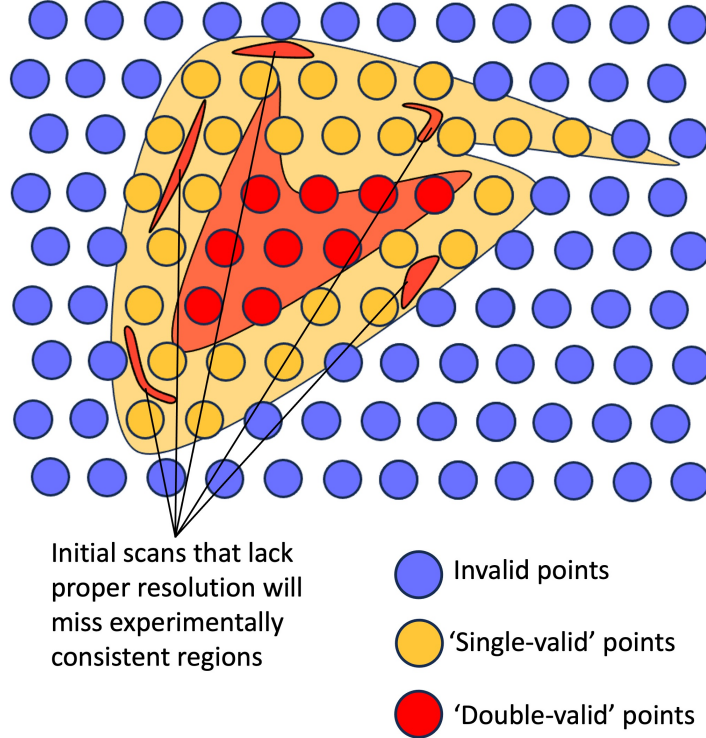


Figure 3.1: A simplified 2D cartoon demonstrating the issue with applying multiple experimental constraints at once on the same dataset/scan. Insufficient resolution means that irregular but desirable regions in theory space may be overlooked. Interesting and unique theories may hide in these enclaves.

3.2 Task & Initial Dataset

To perform an exhaustive search of the pMSSM for points which satisfy multiple experimental constraints, we require a methodology which

- captures the proportionality of the underlying valid subspace with high accuracy,

- allows for the iterative generation of subsequent training datasets, to prevent resolution loss.

In each leg i of the search, we define as ' j -valid' those theory points which meet the $i = j$ -th desired constraint as well as all $j - 1$ previous constraints. Theory points which meet *all* of the listed constraints will be referred to from here on as ' 7 -valid' points. Our list of experimental constraints is defined in Table 3.1 and all references to 'simultaneous constraints' in this work will assume the specific ordering provided therein.

Table 3.1: A list of relevant experimental constraints for our search, presented in order of their application.

j	Constraint	Description
1	$122.09 \text{ GeV} \leq m_h \leq 128.09 \text{ GeV}$ [1, 37]	Mass of Higgs Boson
2	$m_{LSP} = \chi_1^0$ [104]	Neutralino LSP
3	$0.08 \leq \sigma_{DM} h^2 \leq 0.14$ [122, 17]	Dark Matter Relic Density
4	$5.6 \times 10^{-10} \leq a_\mu^{SUSY} \leq 44.6 \times 10^{-10}$ [18, 5]	Magnetic Muon Moment
5	$0.7 \leq \frac{BR(\bar{B}_u \rightarrow \tau \nu)}{BR(\bar{B}_u \rightarrow \tau \nu)_{SM}} \leq 1.38$ [7, 70]	Ratio of MSSM/SM Branching Fractions
6	$2.142 \times 10^{-9} \leq BR(B_s^0 \rightarrow \mu^+ \mu^-) \leq 3.658 \times 10^{-9}$ [83]	Dimuon Branching fraction
7	$3.104 \times 10^{-4} \leq BR(B \rightarrow X_s \gamma) \leq 3.756 \times 10^{-4}$ [70]	Radiative Branching Fraction

We begin with an initial uniform scan of $\mathcal{O}(20.9M)$ pMSSM points, as described in Table 3.2. The subset of this dataset which satisfies the $j = 1$ constraint, the mass of the Higgs Boson, is then used to train our first flow. All theory points are evaluated via SOFTSUSY 4.1.10 and MicrOMEGAS 5.2.6 [8, 16]. We normalize all data values to between 0 and 1 before training.

3.3 Approach

3.3.1 A Brief Introduction to Normalizing Flows

Normalizing flows are invertible mappings which take points from an easy-to-sample base distribution, typically a gaussian, to a more complex target distribution [88]. They are known for high fidelity and low bias in recovering the proportionality of the desired distribution [110,

Table 3.2: Ranges for the pMSSM parameters used in the initial brute-force scan D_0 . Before training an additional pre-processing step was taken to manually remove the gaps in $|M_1| < 0.05$ TeV, $|M_2| < 0.1$ TeV, and $|\mu| < 0.1$ TeV. This step was reversed during the generation of new data.

Parameter	Domain	Description
$ M_1 $	[0.05,4] TeV	Bino Mass
$ M_2 $	[0.1,4] TeV	Wino Mass
M_3	[0.4,4] TeV	Gluino Mass
$ \mu $	[0.1,4] TeV	Bilinear Higgs Mass
M_A	[0.1,4] TeV	Pseudo-scalar Higgs Mass
$ A_b $	[0,4] TeV	Trilinear Bottom Coupling
$ A_t $	[0,4] TeV	Trilinear Top Coupling
$ A_\tau $	[0,4] TeV	Trilinear τ Coupling
$m_{\tilde{L}_1}$	[0.1,4] TeV	1st Generation L.h. Slepton Mass
$m_{\tilde{L}_3}$	[0.1,4] TeV	3rd Generation L.h. Slepton Mass
$m_{\tilde{e}_1}$	[0.1,4] TeV	1st Generation R.h. Slepton Mass
$m_{\tilde{e}_3}$	[0.1,4] TeV	3rd Generation R.h. Slepton Mass
$m_{\tilde{Q}_1}$	[0.4,4] TeV	1st Generation L.h. Squark Mass
$m_{\tilde{Q}_3}$	[0.2,4] TeV	3rd Generation L.h. Squark Mass
$m_{\tilde{u}_1}$	[0.4,4] TeV	1st Generation R.h. up-type Squark Mass
$m_{\tilde{u}_3}$	[0.2,4] TeV	3rd Generation R.h. up-type Squark Mass
$m_{\tilde{d}_1}$	[0.4,4] TeV	1st Generation R.h. down-type Squark Mass
$m_{\tilde{d}_3}$	[0.2,4] TeV	3rd Generation R.h. down-type Squark Mass
$\tan\beta$	[1,60]	Ratio of Higgs VEVs

67, 87, 99]: a crucial property for the iterative nature of our task, to minimize compounding of errors. This mapping is learned by maximizing the log-likelihood of training data from the desired distribution [88]. For our purposes, we choose a map which is the composite of successive autoregressive transformations, specifically rational quadratic splines [110, 52] defined over the interval $[-B, B]$. A more rigorous description of this methodology can be found here: [77, 20] .

3.3.2 Iterating Our Flows

Our iteration procedure is as follows:

1. Flow F_i is trained on data T_i to generate points which satisfy i constraints simultane-

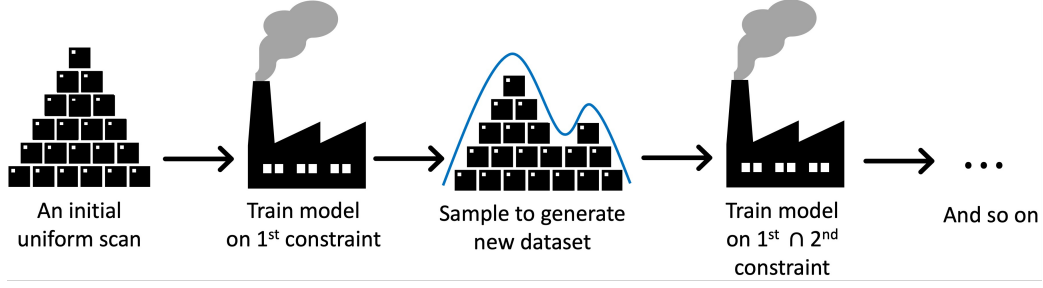


Figure 3.2: A simplified cartoon demonstrating an iterative training regimen. While our methodology uses the cumulative valid points from all prior datasets to train the next flow, in practice the majority of these points frequently come from the most recent preceding flow.

ously, using the constraint ordering listed in Table 3.1.

2. A new dataset D_i is generated from F_i and evaluated.
3. If O_{in} is the subset of D_i which satisfies *only* the n th constraint,² then we can define

$$S_i^j = \bigcap_{n=1}^j O_{in},$$

as the subset of D_i which satisfies j constraints *simultaneously*, using the Table 3.1 ordering, and the efficiency of the i th leg of the search under j simultaneous constraints can be expressed as $E_i^j = \frac{|S_i^j|}{|D_i|}$.

4. Our new training data can then be written as:

$$T_{i+1} = \bigcup_{n=1}^i S_n^{i+1},$$

where D_0 represents our initial scan as described in Table 3.2. For example, $T_2 = S_0^2 \cup S_1^2$.

5. Repeat the process, training F_{i+1} on T_{i+1} , and generating D_{i+1} , and so on.

²While it is possible, and in fact desirable, for individual points in O_{in} to satisfy other constraints beyond the n th constraint, there is no guarantee of this, and the set is defined as only those points from D_i which meet the n th constraint.

To ensure our search resolution is not lost, each new D_i generated consists of $\mathcal{O}(22M)$ points. We assert that this high number of evaluations allows us to search the space more thoroughly than contemporary ‘as-you-go’ approaches, while applying more constraints than standard proportional searches.

3.4 Results and Discussion

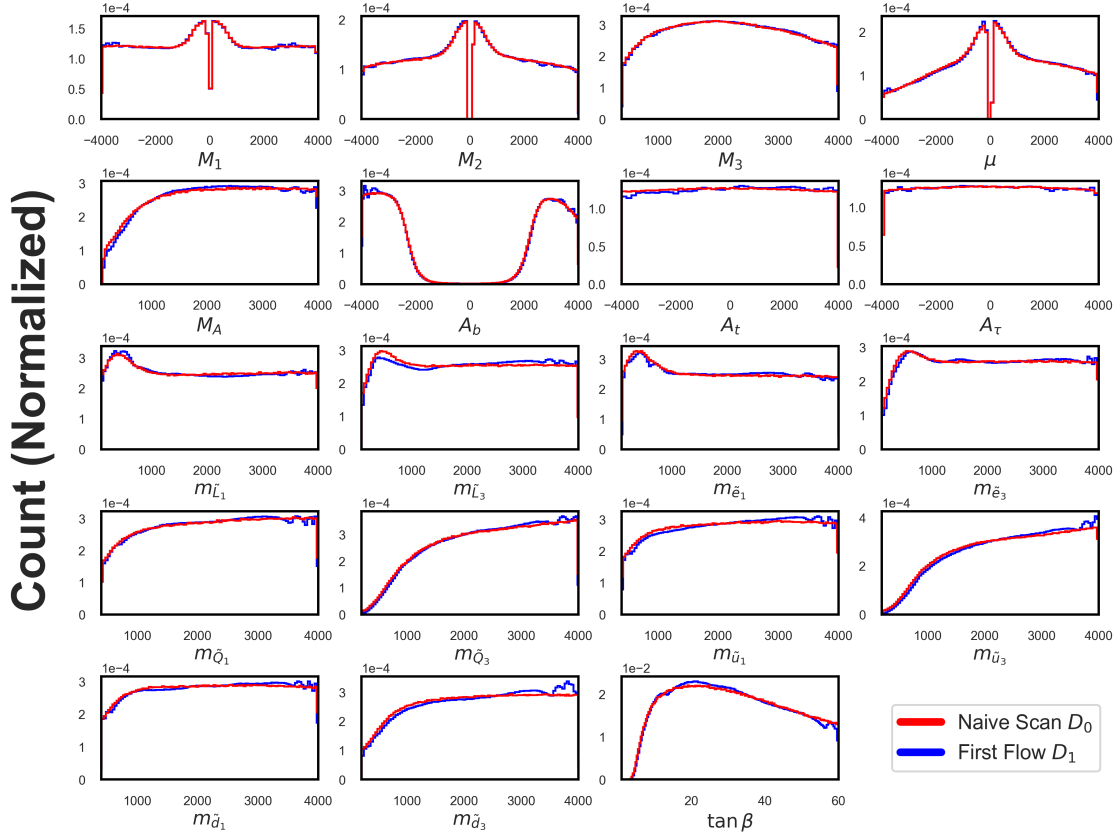


Figure 3.3: The normalized distributions of Higgs-valid pMSSM subsets are shown here: S_0^1 from the naive scan D_0 in red and S_1^1 from the flow-generated D_1 in blue. The proportions are a close match, demonstrating that our flow is behaving as expected in capturing the proportionality of the Higgs-valid subspace.

Overall we find our search method to be highly effective, generating points which meet all 7 experimental constraints simultaneously and with high efficiency. Looking at the results of

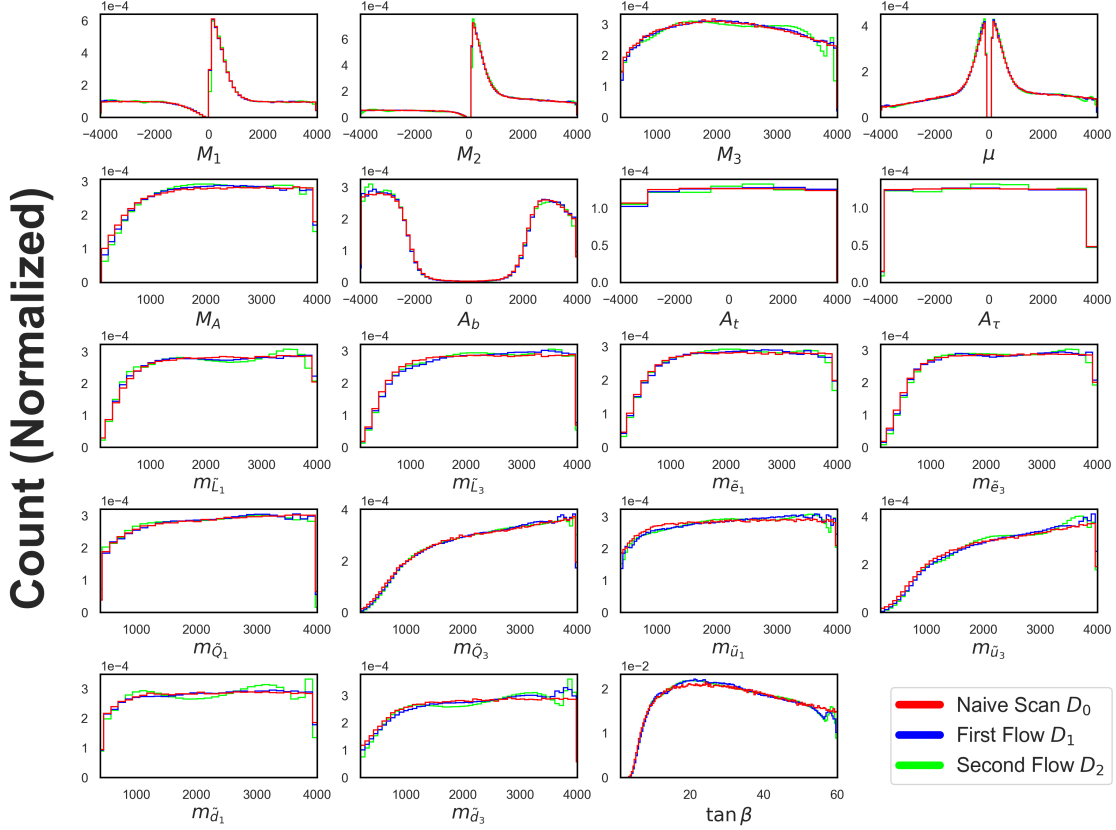


Figure 3.4: The normalized distributions of ‘2-valid’ pMSSM subsets are shown here: S_0^2 from the naive scan D_0 in red, S_1^2 from D_1 in blue, and S_2^2 from D_2 in green, whose flow was trained on $S_0^2 \cup S_1^2$. We observe only minor deviations from the expected proportionality.

the first two legs of the training in figs. 3.3 and 3.4, we demonstrate that the flow captures the proportionality of the Higgs-valid subspace accurately and can then be iterated to the next constraint without introducing significant bias. Over the course of iterating, with only $\mathcal{O}(174.9M)$ cumulative evaluations we are able to generate a total of $\mathcal{O}(43.7M)$ 7-valid points: a brute force scan would require $\mathcal{O}(12.8T)$ evaluations to produce the same quantity, making the task essentially intractable. The resulting end state normalizing flow yields an overall efficiency improvement of more than 5 orders of magnitude compared to a naive scan.

Furthermore, we find specifically that the iterative nature of our approach allowed for a level of high-resolution sampling that would not have been possible otherwise. Our naive scan had only 71 points which met all 7 constraints simultaneously: a single flow trained from

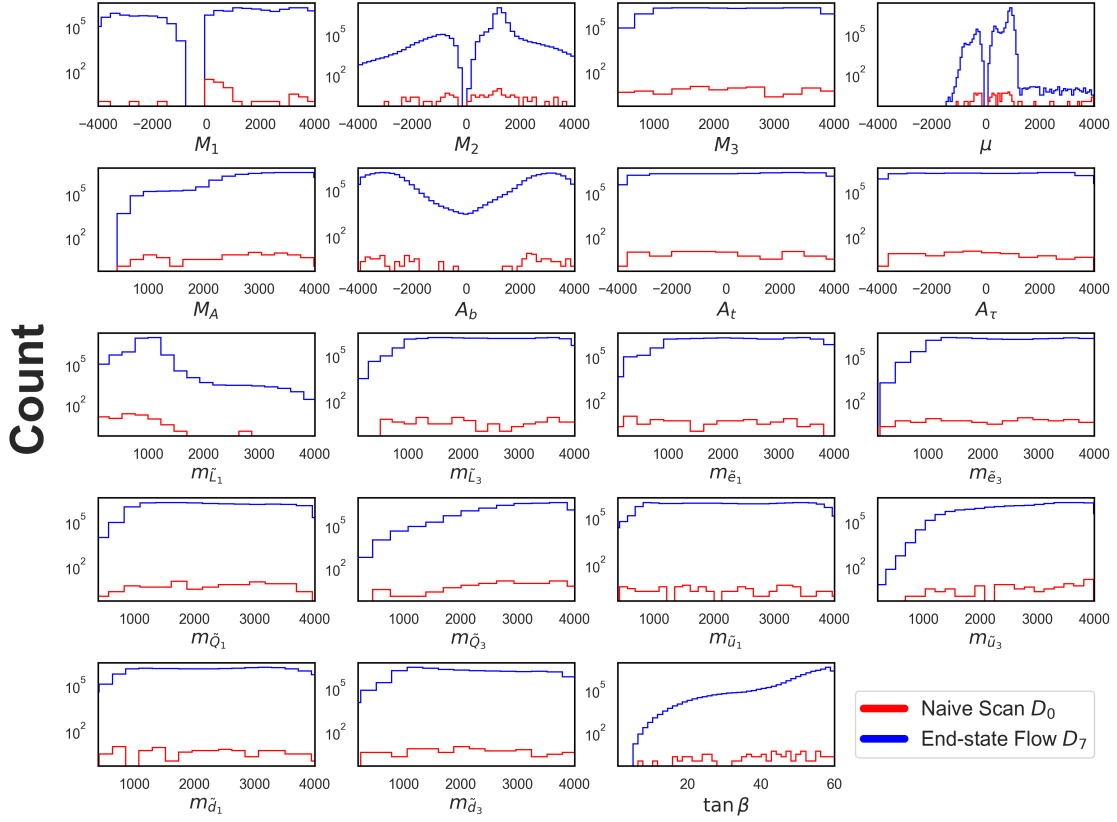


Figure 3.5: The distribution of 7-valid pMSSM points S_0^7 from the naive scan D_0 are shown in red, against the 7-valid points S_7^7 from the flow-generated D_7 which are shown in blue. Several regions where the initial scan found few or no 7-valid points are in fact still viable. See discussion below for specific ranges.

only those points would undoubtedly miss subtle details of the valid manifold where novel theories might hide.

Examining figure 3.5, we see that our end-state flow was able to sample 7-valid points in several regimes where the initial scan lacked sufficient resolution, most notably $M_1 < 0$ GeV, $|A_b| < 2000$ GeV and $m_{\tilde{L}_1} > 1500$ GeV, but also the regions $M_2 < 2500$ GeV, $1000 \text{ GeV} < \mu < 3500$ GeV, and $\tan \beta < 15$.

Looking at the weak-scale observable space (Fig. 3.6) we see similarly that the naive scan had significant difficulty finding 7-valid points in multiple regions, most visibly in the $m_h > 125.5$ GeV, $m_{LSP} > 900$ GeV, $a_\mu^{SUSY} > 2.0$, $\frac{BR(B_u \rightarrow \tau \nu)}{BR(B_u \rightarrow \tau \nu)_{SM}} < 0.9$ and $2.2 < BR(B_s^0 \rightarrow \mu^+ \mu^-) < 2.6$ sectors

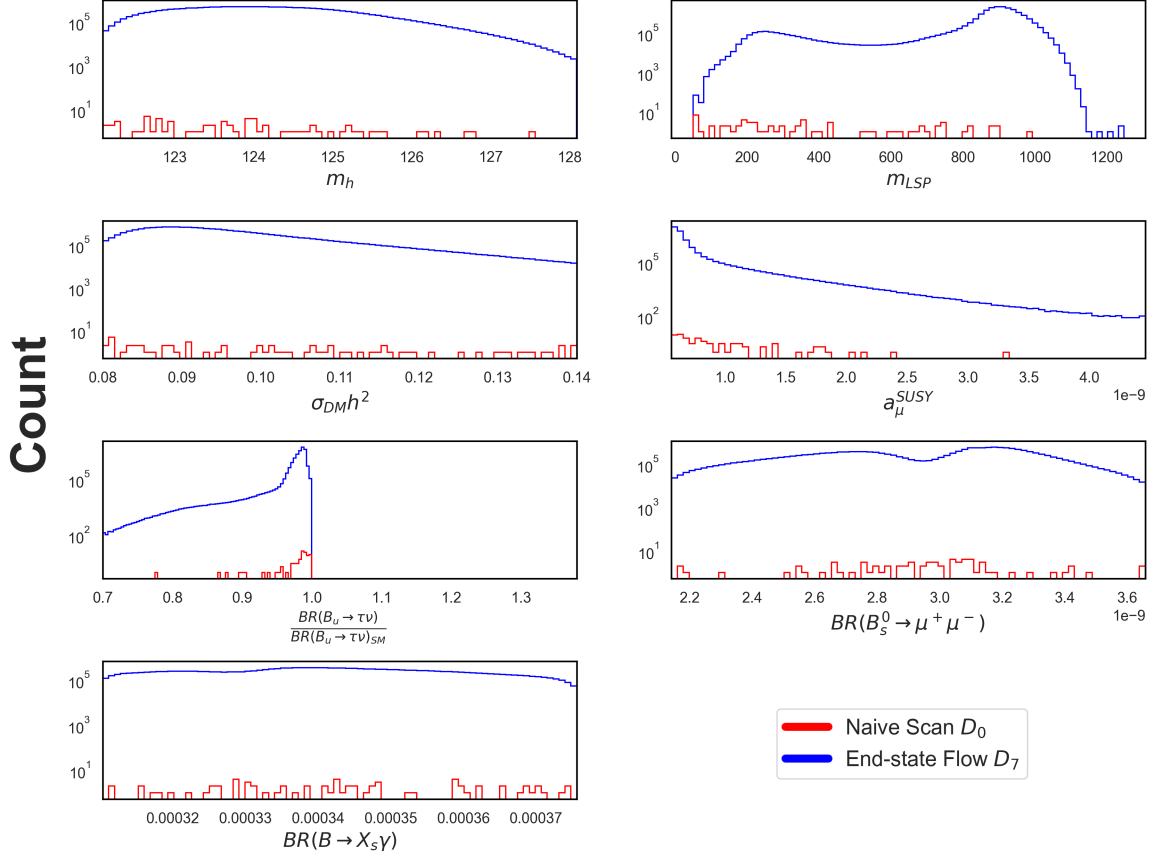


Figure 3.6: The weak-scale observables of the two 7-valid subsets shown in figure 3.5 are shown here, as distributions in ‘constraint space’. S_0^7 from D_0 is shown in red, and S_7^7 from D_7 in blue. Note that since constraint 2 ($m_{LSP} = \chi_1^0$) is not a range of values, the actual value of m_{LSP} for points which meet this constraint is plotted here instead.

of the valid subspace. Our normalizing flow was able to generate points from these regions without difficulty.

Looking at the 2D histograms in figs. 3.7 and 3.8, we see prominent islands/peninsulas in the ranges $600 \text{ GeV} \lesssim m_{LSP} \lesssim 1100 \text{ GeV}$ and $m_{\tilde{L}_1} \gtrsim 1500 \text{ GeV}$ which the naive scan largely overlooked. Proceeding iteratively allowed us to ‘dig deeper’ into the space and sample areas the naive scan could not capture. We also quantify this in fig. 3.9, by calculating the distance from each flow-generated 7-valid point to the nearest naive 7-valid point. We plot this alongside the distance from each naive 7-valid point to the nearest non-identical naive 7-valid point, as a baseline.

Table 3.3: Table of efficiencies for each flow and set of constraints, where a "valid point" in each row is defined as one that meets both a.) the specified constraint, and b.) all prior constraints. Each dataset consists of $\mathcal{O}(22M)$ theory points with the exception of D_0 which contains $\mathcal{O}(20.9M)$ points. The efficiency term for the i th leg of the training under j cumulative constraints, following the ordering in table 3.1, can be written as $E_i^j = \frac{|S_i^j|}{|D_i|}$. The $E_i^{j=i}$ terms are marked here in bold.

j	Constraints (cumulative)	Dataset							
		$D_{i=0}$ (Naive scan)	D_1	D_2	D_3	D_4	D_5	D_6	D_7
1	m_h	0.17081	0.87865	0.86881	0.74879	0.82514	0.91844	0.93258	0.93844
2	m_{LSP}	0.04407	0.23042	0.83470	0.67529	0.71766	0.89695	0.92749	0.93365
3	$\sigma_{DM}h^2$	0.00115	0.00576	0.02025	0.44854	0.21395	0.72421	0.83947	0.88689
4	a_μ^{SU5Y}	1.0×10^{-5}	5.1×10^{-5}	0.00016	0.00024	0.18960	0.70604	0.82870	0.87775
5	$\frac{BR(B_u \rightarrow \tau \nu)}{BR(B_u \rightarrow \tau \nu)_{SM}}$	9.0×10^{-6}	4.5×10^{-5}	0.00014	0.00023	0.17998	0.70459	0.82850	0.87767
6	$BR(B_s^0 \rightarrow \mu^+ \mu^-)$	6.3×10^{-6}	3.2×10^{-5}	9.8×10^{-5}	0.00017	0.14818	0.59905	0.81750	0.87307
7	$BR(B \rightarrow X_s \gamma)$	3.4×10^{-6}	1.9×10^{-5}	5.9×10^{-5}	0.00011	0.10178	0.42509	0.59386	0.86360

Our 7-valid points also reproduce known physics structures, for instance: the Bino line [53] is visible in Fig. 3.7c, the Higgsino line [69] in Fig. A.4e, the A-funnel [60] in Fig. 3.7e, and the asymptotic behavior of $\frac{BR(B_u \rightarrow \tau \nu)}{BR(B_u \rightarrow \tau \nu)_{SM}}$ at large M_A due to charged-Higgs suppression [81] in Fig. A.4d.

Additionally, across all legs of our search we find no 7-valid points in the regimes of $\mu < -2600$ GeV in the GUT-scale parameter space and $\frac{BR(B_u \rightarrow \tau \nu)}{BR(B_u \rightarrow \tau \nu)_{SM}} > 1.0$ in weak-scale parameter space. The naive scan did probe these regions, with $\mathcal{O}(3.75M)$ and $\mathcal{O}(21.8k)$ pMSSM points respectively, but cumulatively our flows were able to generate an additional $\mathcal{O}(5.52M)$ and $\mathcal{O}(129.3k)$ points in each of these regimes, none of which were 7-valid. Additionally, our flows found points approaching these boundary regions which the naive scan missed: the minimum 7-valid value for μ in the naive scan was -1041.61 GeV, but our T7 set contains individual '7-valid' points with values as low as $\mu = -2530.06$ GeV. Similarly, the highest value for $\frac{BR(B_u \rightarrow \tau \nu)}{BR(B_u \rightarrow \tau \nu)_{SM}}$ amongst the naive 7-valids was 0.99946, but in T7 we found values as high as $\frac{BR(B_u \rightarrow \tau \nu)}{BR(B_u \rightarrow \tau \nu)_{SM}} = 0.99991$. Our results suggest the valid density in these regions can be constrained to a lower threshold than the naive scan would allow. Our improved search

resolution allows us to make stronger inferences about which portions of the space might be ruled out.

3.4.1 Improvements in Efficiency

As expected, we see that after each stage of training there are significant improvements in the $E_i^{j=i}$ term (the efficiency of D_i under $j = i$ simultaneous constraints) as seen in bold in table 3.3. This confirms our flow is learning to generate the desired distributions accurately at each step. In addition to this, we observe varying, though usually incremental improvements in the $E_i^{j>i}$ efficiencies at each stage: for instance the E_1^2 efficiency outperforms the E_0^2 efficiency. We attribute these correlations in efficiency to the underlying structure of the space and the spatial overlap of the various O_{ij} subsets. The efficiencies in the $E_i^{j<i}$ regime are much less straightforward.

For $i \leq 3$ we observe that these terms *decrease* in efficiency at each step. Naively we posit that this represents the model focusing on the constraint at hand at the expense of previous learned constraints. However, at $i = 4$ we see this trend begin to reverse, such that the $E_i^{j<i}$ efficiencies are now improving after each training. Stranger still, this change does not happen uniformly across all efficiencies until $i = 5$. While E_4^1 and E_4^2 are improvements over their D_3 counterparts, the efficiency for E_4^3 still decreases in comparison to E_3^3 . A partial explanation may be that the iterative process allows the flows to transition from initial "exploration" at the boundaries of these overlapping subsets into later "exploitation" of the regimes it has found. For instance, perhaps the $j = 3$ or $j = 4$ constraint create particularly irregular or disjointed manifolds, but the subsets of those manifolds which are valid under the $j = 5$ constraint become neater and easier to sample, either because of the geometry of the space or because our iterative method does not lose resolution as we 'zoom in'.

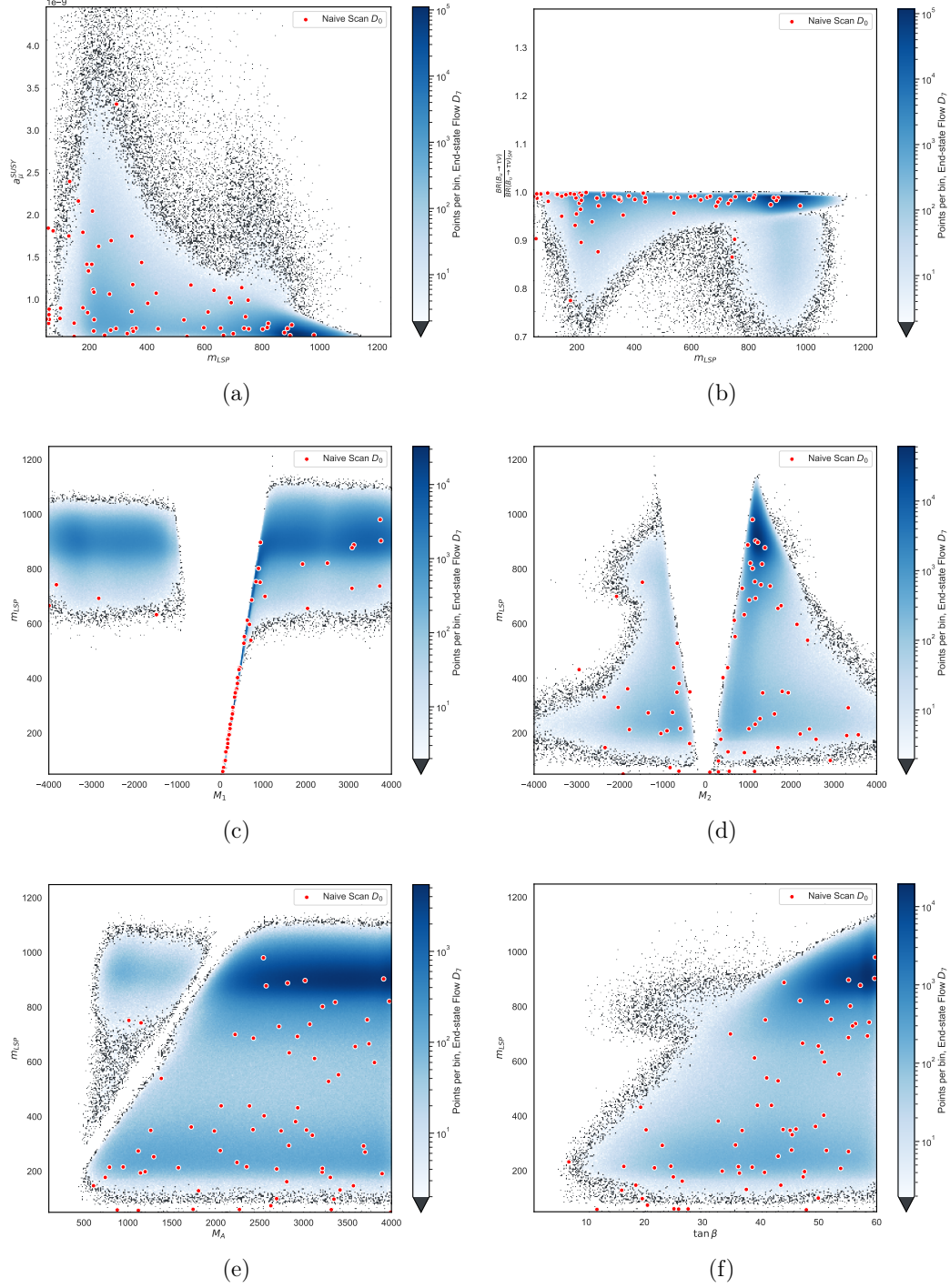
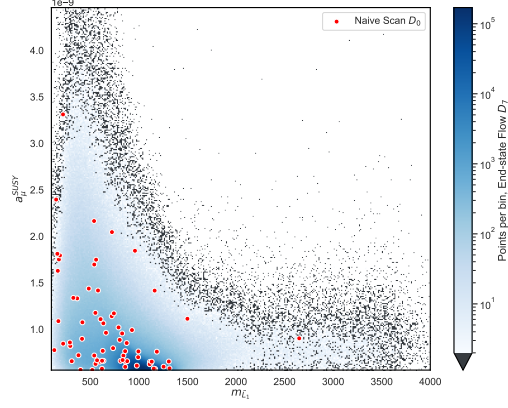
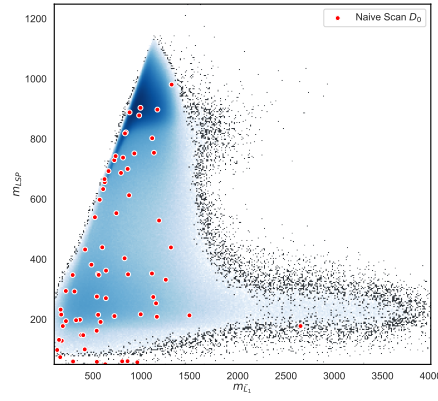


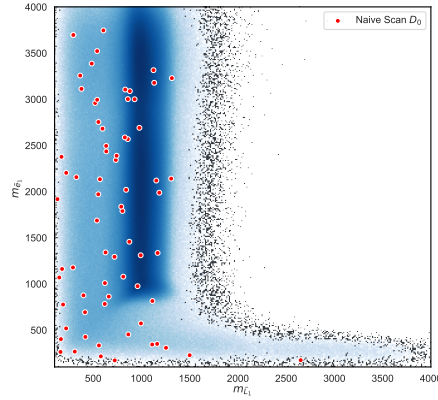
Figure 3.7: Selected 2D histograms comparing the 7-valid distributions S_0^7 from D_0 in red, and S_7^7 from D_7 in blue. Bins containing only a single theory point from S_7^7 are shown in charcoal for improved visibility in sparse regions and the axes for GUT-scale parameters are set to exclude points outside the initial scan boundaries. Note the islands/peninsulas found around $600 \text{ GeV} \lesssim m_{LSP} \lesssim 1100 \text{ GeV}$ which the naive scan largely missed, but which the flow was able to sample admirably. The parameters plotted are as follows: a.) a_μ^{SUSY} vs. m_{LSP} b.) $\frac{BR(B_u \rightarrow \tau \nu)}{BR(B_u \rightarrow \tau \nu)_{SM}}$ vs. m_{LSP} c.) m_{LSP} vs. M_1 d.) m_{LSP} vs. M_2 e.) m_{LSP} vs. M_A f.) m_{LSP} vs. $\tan \beta$



(a)



(b)



(c)

Figure 3.8: Selected 2D histograms comparing the 7-valid distributions S_0^7 from D_0 in red, and S_7^7 from D_7 in blue. Bins containing only a single theory point from S_7^7 are shown in charcoal for improved visibility in sparse regions, and the axes for GUT-scale parameters are set to exclude points outside the initial scan boundaries. Note the islands/peninsulas found around $m_{\tilde{L}_1} \gtrsim 1500$ GeV which the naive scan largely missed, but which the flow was able to sample admirably. The parameters plotted are as follows: a.) a_μ^{SUSY} vs. $m_{\tilde{L}_1}$ b.) m_{LSP} vs. $m_{\tilde{L}_1}$ c.) $m_{\tilde{e}_1}$ vs. $m_{\tilde{L}_1}$

3.4.2 Comparability to other methods

Direct comparison to other state-of-the-art methods proves difficult: Differences in evaluation software, constraint quantity and window size, and even the number of free vs. fixed parameters make direct juxtaposition difficult [13].

Adding to this, most proportional methods focus on maximizing the efficiency of generated points across only one or two constraints, perhaps due to low resolution, while the "as-you-go" methods that can more robustly tackle many constraints at once tend to emphasize lowering efficiency primarily through minimizing the number of software evaluations instead of maximizing a percent yield.

Nevertheless, in Figure 3.10 we plot a broad strokes comparison of efficiency vs. number of experimental constraints, across a number of recent search methodologies in the pMSSM. While this does not account for the individual difficulty of each experimental constraint, it is notable perhaps that of the proportional studies, we found no other methods capable of generating theory points which are valid for > 2 constraints simultaneously, and that for the 'as-you-go' methods the highest efficiency yield was only 0.05, more than an order of magnitude lower than our E_7^7 efficiency of 0.86360.

We also note here that efficiency is not a legitimate measure of the breadth or thoroughness of a search³. Most pMSSM searches do not include any direct measure which quantifies this 'proportionality fidelity,' and those which do [44] present their measurements separately for individual theory parameter, as opposed to using a single 'global' metric, which makes cross-study comparisons a challenge. While we have suggested in the past [13] that Sliced Wasserstein distance [92, 78, 91] could serve as a metric in this regard, even this would not

³In the most extreme cases, such as mode collapse [93], the repeated generation of the same individual valid point regardless of input data would give a misleading efficiency of 100%. One possible solution is to only consider the number of unique generated points, but this causes complications in the subjective determination of what is 'sufficiently unique,' and does not address the underlying issue of quantifying how accurately a given search captures the true valid distribution within the space.

address the issue we pose in footnote 1: that discovering new or unexpected regions which are unrepresented in the initial naive scan would result in a worse score with such a metric.

The dearth of methods which can do high-efficiency proportional point generation across multiple constraints further underlines the necessity of iterative approaches which can bridge this gap in order to generate points that are valid across many constraints with high efficiency, and do so while providing more exhaustive coverage of the space than the ‘as-you-go’ methodologies.

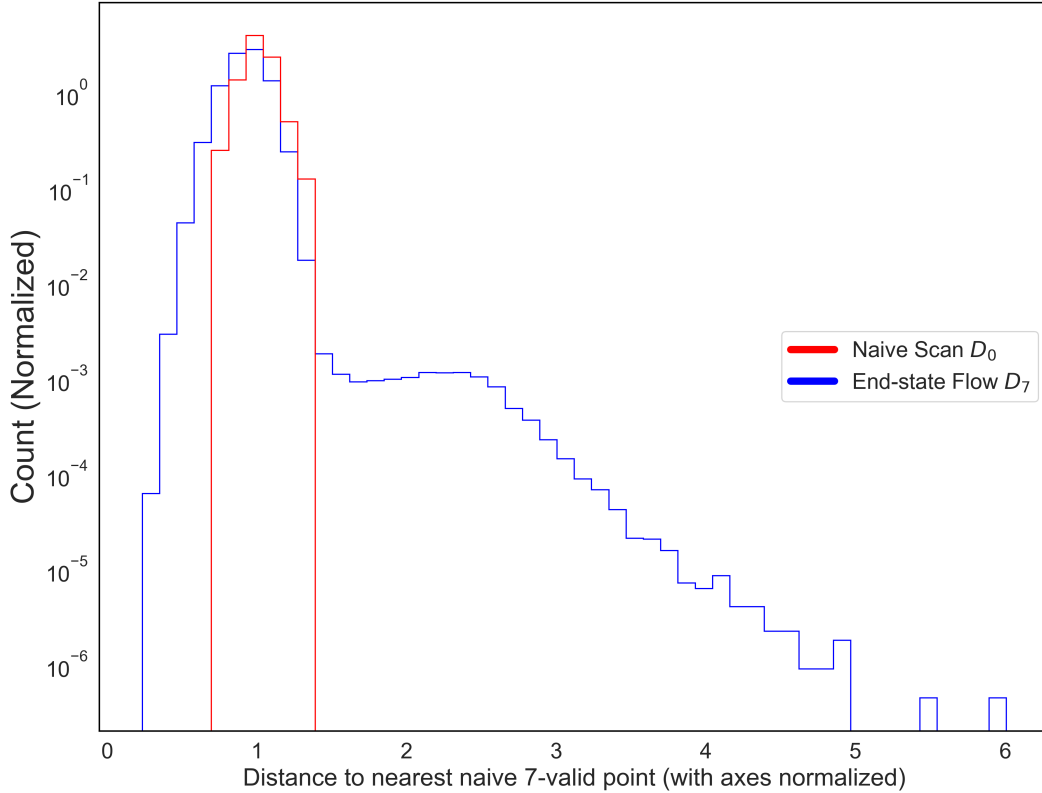


Figure 3.9: The distribution of distances from each point in S_0^7 in red and S_7^7 in blue to the nearest non-identical neighboring point in S_0^7 . Distances were calculated with axes normalized using the values in Table 3.2. The ranges of these distributions suggest our flow was able to discover new 7-valid points that the naive scan missed due to low resolution (i.e. distances $\lesssim 1.5$) as well as those outside the range of the initial scan (See Appendix A.3.1)

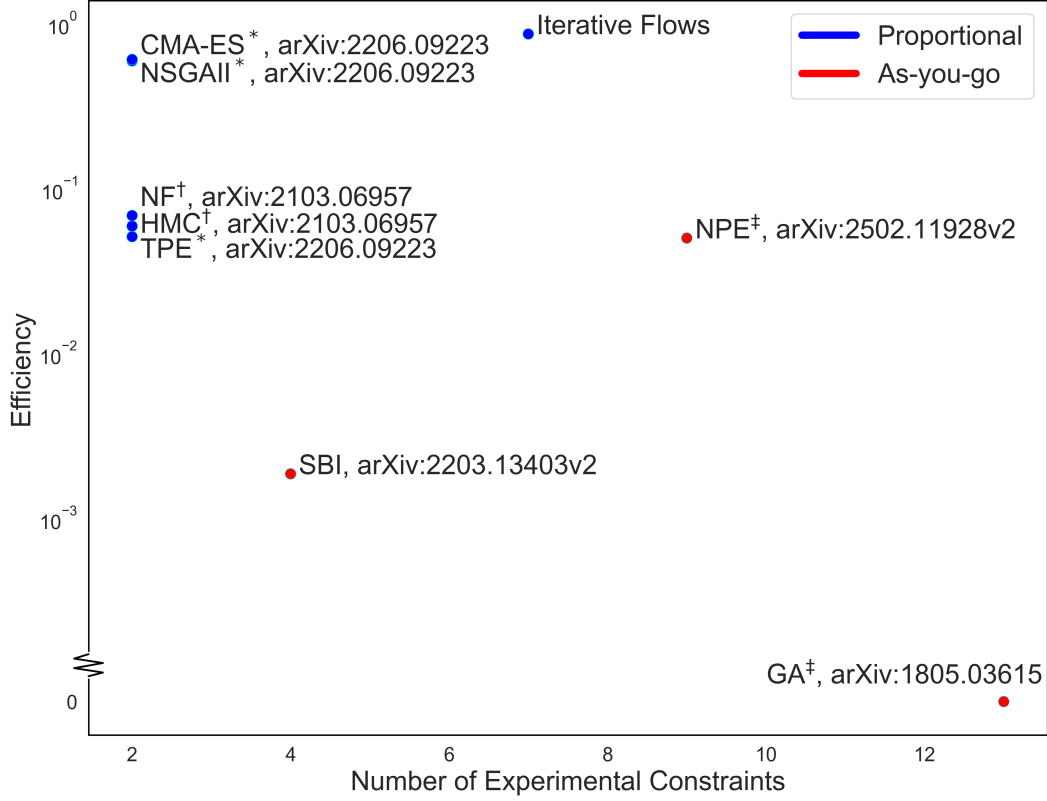


Figure 3.10: A scatterplot of recent SOTA pMSSM search methods showing efficiency vs. number of experimental constraints. The studies cited are as follows: CMA-ES [44], NSGAII [44], TPE [44], HMC [77], Normalizing Flows [77], SBI [104], NPE (amortized SBI) [38] and Genetic Algorithms [3] are plotted against our results (listed here as "Iterative Flows"). Many of the proportional search papers listed here also conducted single-constraint trials (typically using the Higgs Mass constraint) which achieved higher efficiencies than their respective two-constraint runs. These single-constraint runs are omitted here for legibility. Please note the following caveats and stipulations:

* Efficiency values for [44] are taken from [*Phys. Rev. D* 107, 035004 (2023)], which differ from the values specified in the arxiv preprint.

† Efficiency values for [77] are taken from [*Eur. Phys. J. C* 81, 1138 (2021)] which differ from the values specified in the arxiv preprint. The evaluation software used in [77] has known computational differences which may have led to artificially lowered efficiencies, as noted in Footnote 10 of [44].

‡ The convention used in [3] contains 'aggregated composite' constraints such as HiggsSignals [15] which combines multiple signals into a single experimental constraint with a single χ^2 contribution. We count each such composite as one individual constraint, and as such, when observables which appear within that composite block are instead studied individually, as in [38], we group and count those as one individual constraint as well. This discrepancy in how individual constraints are denoted is illustrative of the broader difficulties in comparing results across multiple pMSSM search studies. We also note that [3] contains two separate trials with nearly identical constraints, with the exception that δa_μ^{SUSY} in trial 1 is replaced with the GCE constraint [35] in trial 2 (in both cases, no theory points which meet all 13 constraints were found).

3.5 Conclusion

We trained a series of normalizing flows iteratively on 19D pMSSM distributions of experimental constraints, using the output from each to constitute the training data of the next, resulting in an end-state flow which can probe a total of 7 experimental constraints at once, without significant loss of resolution. Our results demonstrated a significant $\mathcal{O}(10^5)$ improvement over naive efficiencies as well as an improvement in resolution over non-iterative methods such as brute-force scans.

However one possible drawback of this iterative framework is that if additional observables are desired which were not included in the original MicrOMEGAS evaluation script, significant re-evaluations would be required to calculate the desired observable for all prior datasets.

Another potential drawback is the potential under-sampling of sparse but viable regions: Because sparse regions of the valid manifold are harder to sample than dense ones, the flow has fewer points to train on in the areas where the constraints are already the most challenging. When iterating, this bias has the potential to compound, causing the model to over-sample dense regions which are easier to learn at the expense of more isolated and potentially interesting ones. However, even in this worst case scenario, the model will still capture some aspect of the proportionality more reliably than 'as-you-go' methods, which do not explore the entire search space and tend to hone in on the first viable region they find.

More work remains in probing the depths of the pMSSM, most notably with regard to potentially challenging constraints such as the Galactic center gamma ray excess constraint (GCE) from Fermi-LAT [3, 35]. Flushing out the isolated, novel, and unexplored regions that hide in high dimensional parameter spaces, and doing so *exhaustively*, will require methodologies that can dig deep across multiple difficult constraints without loss of breadth or data resolution. We posit, optimistically, that iterative methodologies have a role to play

in the future of this work, both in searching for promising 'islands' and ruling out portions of parameter space, in increasingly higher-dimensional HEP applications. Such discoveries have the potential to motivate new LHC analyses for supersymmetric signatures. In the words of Blues Traveler's John Popper: "as we seek so shall we find". When high-dimensionality makes brute-force scans intractable, we will find novel theory-space enclaves only by seeking them out.

Chapter 4

Reversing Spaces: Towards an inverse mapping of the PMSSM

Abstract— We introduce a framework for learning mappings from weak scale observables to corresponding GUT-scale theory parameters in the context of the cMSSM and pMSSM, the reverse of the typical GUT-to-weak mapping done by typical HEP software evaluations. In general this type of inverse of complex Renormalization Group Equations (RGEs) is considered intractable. Our ‘correctional network’ methodology skirts this difficulty by focusing on correcting an initial ‘bad guess’ theory point in GUT space, which also allows us to avoid many degeneracy issues around what can otherwise be a one-to-many mapping, where multiple GUT points have the same weak signature. We also demonstrate the utility of iterative corrections, where the network output is re-added to the original input to generate additional improvements in performance, as well as a secondary ‘Smart iteration’ network which monitors this process and determines when sufficient iterative correction has been achieved. This is an on-going project, so the results presented here are of a preliminary nature.

4.1 Introduction

A fundamental difficulty with brute-force searches through High Energy Physics (HEP) theory spaces is their high-dimensionality. To search an n -dimensional space with m possible values per dimension requires m^n data points, which makes searching spaces like the $\mathcal{O}(100)$ -dimensional Minimal Super-Symmetric Model (MSSM) essentially intractable [100]. Complicating things further, the parameters needed to specify a theory are defined at the Grand Unified Theory (GUT) energy scale, while the experimental observables used to constrain theory space are defined at a separate, weak energy scale. Because the mapping from GUT to weak is highly non-linear,

the distribution of viable theories not excluded by experiment takes the form of a complex, discontinuous 'valid manifold' that sits within the high-dimensional GUT-space. Probing this distribution is challenging.

Crucially, while the mapping from GUT to weak is a calculable process, there is no one-to-one, tractable mapping from weak to GUT.

Interestingly however, we observe that many of these difficulties ameliorate when we treat the set of weak-scale observables as their own search space: Since we select which and how many experimental parameters are of interest in any given search, the dimensionality of the search space is easily containable a priori. The constraints on these weak parameters often take the form of a central value with error bars, which we can take to define the edges of a 'valid box': the complex valid manifold from GUT space can now naively be represented as a uniform distribution whose dimensionality matches the number of constraints we wish to apply ¹.

¹While in practice there is no reason to assume a uniform scan of GUT space will return a uniform distribution in weak observable space, we are nevertheless interested in finding any-and-all points which fall within the valid box, especially those from sparse regions which are atypical and may lead to novel theoretical classifications. Thus a uniform observable space distribution is a reasonable ansatz.

Typically evaluation software [8, 16] is used to calculate the observables for a given GUT-scale theory point. However, if one can learn to *approximate* a 'reverse mapping' of that software, the tractable properties of observable space can be used to probe the less-tractable GUT parameter space itself. Such a mapping has the potential to yield sub-constraint-window level precision, meaning its utility could persist even as constraints tighten in the future.

While appealing in prospect, in practice the direct implementation of such a reverse mapping faces multiple hurdles:

1. The HEP calculations which map from GUT to weak-space involve running coupling calculations via Renormalization Group Equations (RGE) that are mathematical complex and non-trivial to reverse with the required precision.
2. There is no guarantee that any given set of weak-scale observables has a one-to-one mapping to a corresponding GUT-space point. In fact, in cases where the dimensionality of the weak space is less than that of the GUT space, the weak-to-GUT mapping can be one-to-many. Such a reverse mapping would not be able to handle the degeneracies that arise when two GUT points have identical weak-space vectors.
3. Certain observables, such as the Dark Matter Relic Density, are known to be highly volatility: small movements in GUT space can produce orders of magnitude shifts in weak space, which means any mapping with insufficient precision will incur significant inaccuracies.

Such difficulties warrant a fundamental shift in approach. In this paper we introduce the framework for a "Correctional Network," which maps from a 'bad guess' GUT point to the nearest GUT point which has the desired weak-scale properties. We use this correctional framework to learn approximate reverse mappings based on two simultaneous experimental constraints, the Higgs Boson Mass and Dark Matter Relic Density in two subspaces of

the MSSM, the the 5-dimenisonal constrained MSSM (CMSSM) and the 19-dimensional phenomenological MSSM (pMSSM).

Instead of attempting to approximate the computationally-difficult inverse of a differential equation, we instead seek to learn directly how small "nudges" in GUT space effect the values of weak-observables. Because we restrict our search to the *nearest* desirable GUT point with respect to the 'bad guess', the potential for degeneracy is reduced significantly². The output of our network is not the desired GUT point itself, but rather the δ vector in GUT space which takes our 'bad guess' to the desired point. We postulate that this makes the network more sensitive to the small movements in GUT space required to handle volatile constraints. Searching the space for a specific weak signature then becomes a matter of simply calculating the "nudges" of many bad guesses for the same desired observable vector, like throwing a handful of marbles into a precisely built funnel.

Additionally, because the network takes GUT points as inputs, our method allows for 'iterative corrections,' wherein the model output is added to the original point and then passed through the network again, successively. To avoid excessive corrections, we also introduce a secondary network we refer to as 'Smart Iteration' which is trained to determine whether outputs from the correctional network are sufficiently 'nudged' to satisfy difficult constraints such as the dark matter relic density, or whether further correction is needed.

We proceed as follows: In section 2 we outline the task and the datasets used. In section 3 we discuss the specifics of our methodology. In section 4 we present our preliminary results in the cmssm and pmssm. In section 5 we discuss future work which remains for this project.

²While it is technically possible for multiple theory points with identical weak signatures to sit equidistant from the 'bad guess' point, in practice we construct our training dataset as a list of ordered pairs, a de facto one-to-one mapping.

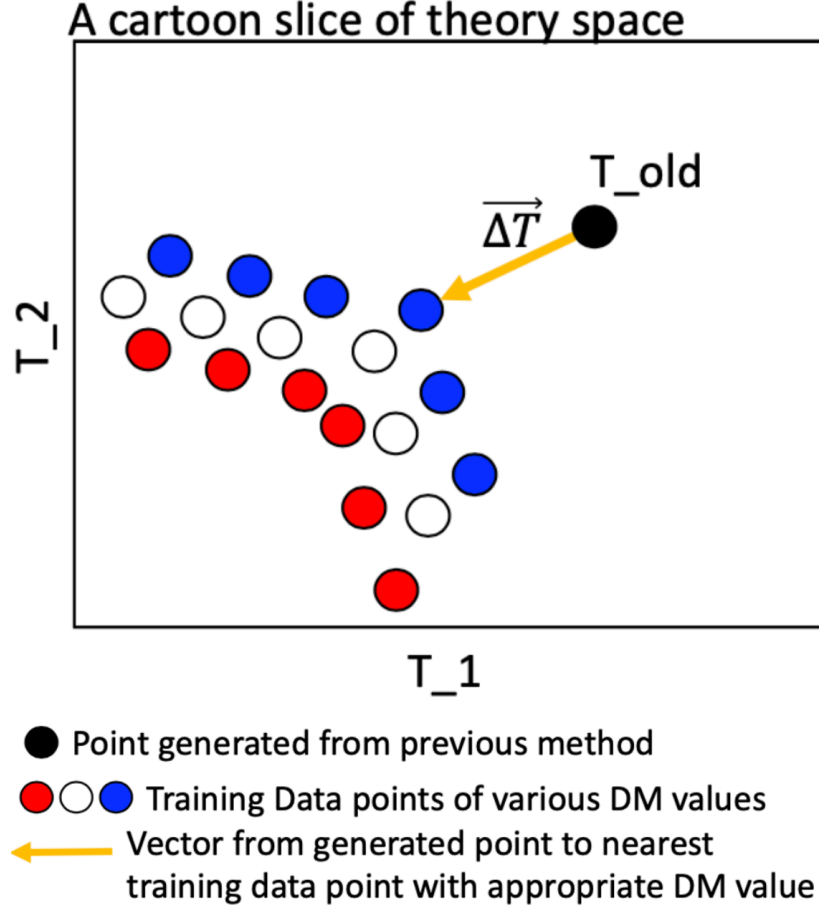


Figure 4.1: A simple cartoon in a 2D slice of theory space demonstrating a desired correction where the vector ΔT maps the proposed T_{old} to the nearest point with the desired weak signature (here dark matter relic density).

4.2 Task & Datasets

4.2.1 Task

To learn an approximate reverse mapping from weak observable space to GUT-scale parameter space, we require a methodology which:

- Avoids degeneracy issues posed by multiple GUT points with the same weak signature
- Avoids the mathematical complexity of directly inverting the RGE calculations.
- Is sensitive to volatile constraints like DM where small shifts in GUT space produce

large shifts in weak space.

We construct our 'valid box' from two experimental constraints, the Higgs boson mass and the dark matter relic density, using the following constraint windows:

- Higgs Boson Mass: $122.09 \text{ GeV} \leq m_h \leq 128.09 \text{ GeV}$ [1, 37]
- Dark Matter Relic Density: $0.08 \leq \Omega_{DM} h^2 \leq 0.14$ [122, 17]

where all evaluations of weak parameters are done via SOFTSUSY 4.1.10 and MicrOMEGAS 5.2.6 [8, 16]. We focus solely on being able to reverse the GUT-to-weak mapping within this valid box: there is little utility in generating theory points that do not agree with experimental constraint.

4.2.2 Correctional Dataset

Our correctional dataset is generated as follows

1. We begin with a library of known valid points (LOKV), which have been generated via past work, including KDE sampling and Normalizing Flows methods [114, 88, 77]. In addition to the two constraints listed above, we require that all valid points in this library have the Neutralino LSP $m_{LSP} = \chi_1^0$. All points in this library, and their corresponding weak values, are normalized to values between 0 and 1.
2. Using KDE sampling, a GUT-scale 'dummy point' is generated based on the distribution of the library of valid points. Desired weak scale values for both constraints are generated via uniform distribution. Together these make up the input vector for the correctional network.
3. The generated weak values are used to filter through the library of known points, such that only those within some tolerance $\pm\sigma_c$ of the specified values are kept.

4. From this culled library, the point which lies closest in GUT space to the 'dummy point' is selected: the difference vector between this target point and our dummy point becomes our target output. We posit that leaving the output in the form of a difference vector, allows our method to be more sensitive to small shifts in GUT space.
5. Repeating this process, we generate a dataset of ordered pairs, where the dummy point and desired weak vector are matched to delta vector which maps to the closest known valid point with the desired weak signature. The full ordered pair can then be written as $([T_{dummy}, W], \Delta T)$, where $T_{dummy} - \Delta T = T_W$ recovers the nearest theory point with the correct weak signature from the LOKV. In practice we generate multiple weak values for each dummy point to encourage more generalized learning

4.2.3 Smart Iteration Dataset

Additionally, we generate a dataset for training the smart iteration network, with the specific goal of distinguishing proposed points by accuracy of dark matter value (i.e. whether the proposed GUT point has the desired DM value). Our procedure is as follows:

1. From each point in our library of known valids, we take the GUT point T_{True} and the Higgs mass m_h . For the true positive class we keep the true dark matter value, such that our order pair takes the form $([T_{True}, m_h, \Omega_{DM}h^2], y = 1)$
2. For the true negative class, we then draw ω_{dummy} from a uniform distribution between 0 and 1, and append this as our proposed DM value, ensuring that it is sufficiently **innacurate** to the specified threshold $\pm\sigma_s$ such that $|\Omega_{DM}h^2 - \omega_{dummy}| > \sigma_s$. The ordered pair then looks like $([T_{True}, m_h, \omega_{dummy}], y = 0)$
3. In practice we generate multiple ω_{dummy} for each T_{True} to improve generalization. We then add duplicates of the positive class datapoints to create an even split between the two classes. We detail the parameters of both the correctional and smart iteration

datasets in table A.4.

4.3 Approach

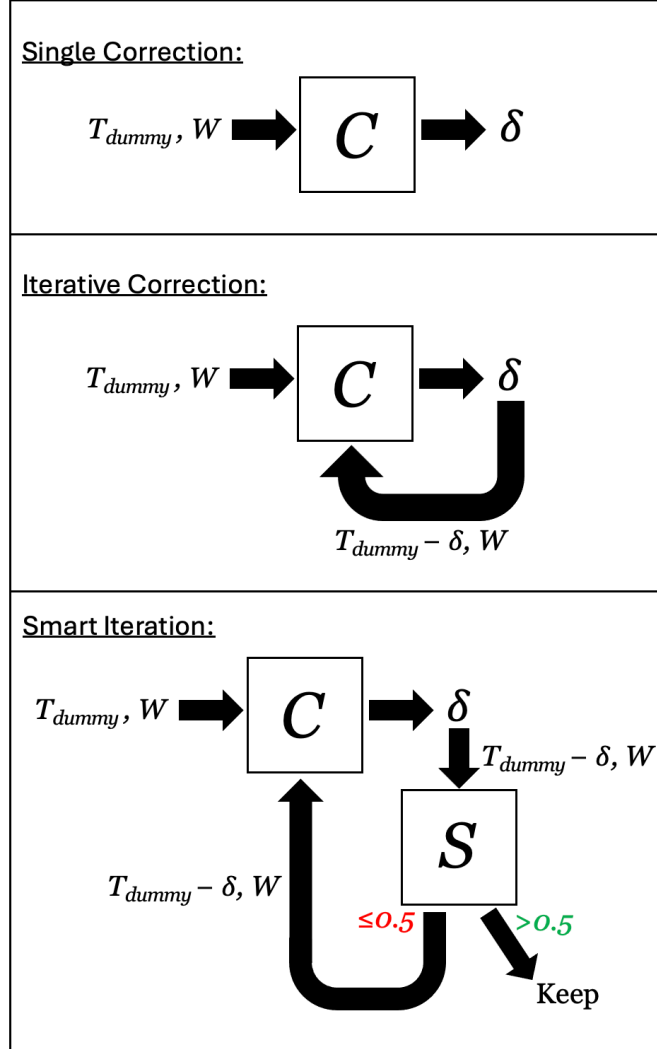


Figure 4.2: A series of diagrams representing the procedures for a single correction, iterative correction, and 'smart iteration' correction, respectively

4.3.1 The Correctional Network

Using pytorch and ADAM optimizer, as well as the following packages [113, 84, 86, 127, 74, 102, 128, 80, 114, 116], we train a neural network C to correct a 'bad guess' theory point T_{dummy} by providing a 'nudge' in theory space such that the resulting point has the desired

weak observable vector W :

$$\delta = C(T_{dummy}, W)$$

Our loss function is then calculated as the MSE loss between the network output and the truth value from the ordered pair:

$$\mathcal{L}_c = \frac{1}{N} \sum_{i=1}^N (\Delta T_i - \delta_i)^2,$$

where N is the batch size. To generate new points, we again use KDE sampling to select dummy points in the vicinity of the LOKV distribution, and append our desired weak inputs. Our network then maps this vector to a GUT-space nudge, and the new 'corrected' theory points can thus be written as

$$T_{new} = T_{dummy} - \delta.$$

4.3.2 Iterative Corrections

Because our method includes a GUT point as an input, iterative corrections are possible. Our output is used to calculate the new GUT point, while the weak input is kept constant:

$$\delta^{j+1} = C(T_{dummy}^j - \delta^j, W),$$

where $T_{dummy}^j = T_{dummy}^{j-1} - \delta^{j-1} = T_{dummy}^0 - \sum_{k=0}^j \delta^k$, and the initial dummy point T_{dummy}^0 is generated via KDE sampling as specified above. In practice we find that the magnitude of the nudge, $\|\delta^j\|$ shrinks with each additional correction, suggesting that the network is closing in on one specific value rather than merely fluctuating between multiple possible candidates chaotically (See fig 4.3).

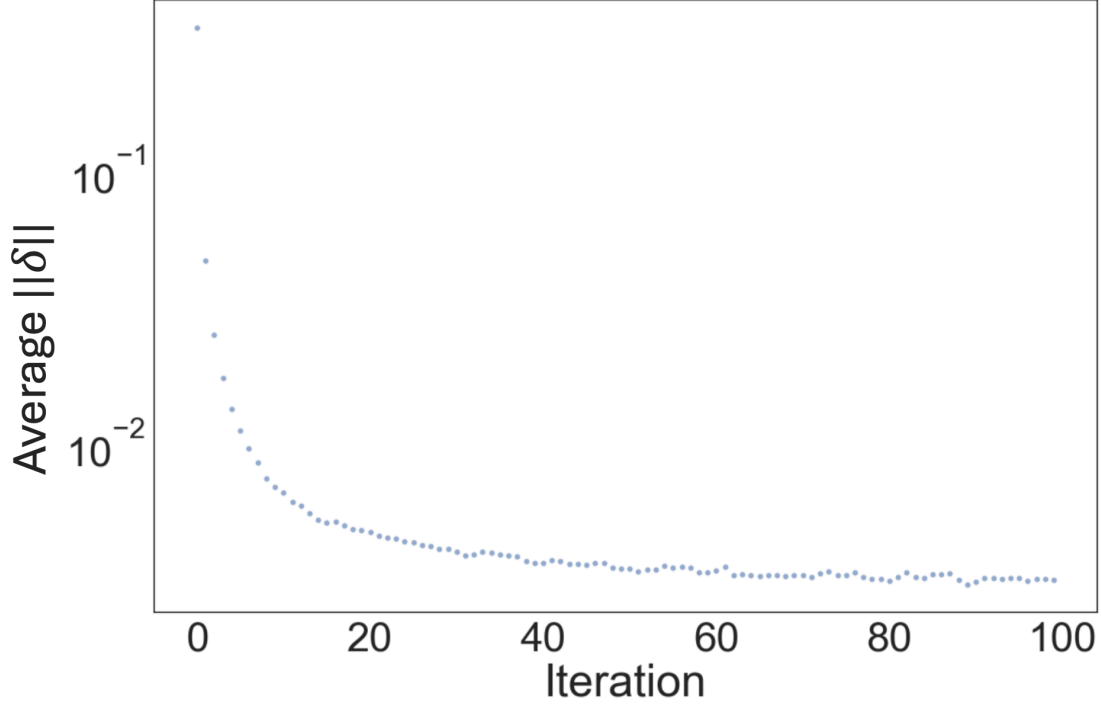


Figure 4.3: The average length of the correction vector $\|\delta\|$ is plotted over 100 iterative corrections for a batch of 10k cMSSM points. We see approximately asymptotic behavior, suggesting that the network is getting closer to the desired solution with each iteration, not vacillating wildly between many possible options.

4.3.3 Smart Iteration

In practice, we find the dark matter constraint significantly more challenging to reverse than the higgs mass, in both the cMSSM and pMSSM. Thus we introduce a secondary network whose job is to monitor whether the implemented correction has had the desired effect. The smart iteration network takes similar inputs to the correctional network but with a significantly simpler task. The neural net $S(T, W)$ is trained to predict whether the input theory point T has the desired weak signature W , producing a single confidence value $S(T, W) = \hat{y}$ between 0 and 1. We then utilize the MSE loss function to measure how well the network reproduces the true value y from the ordered pair:

$$\mathcal{L}_s = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2,$$

where N is the batch size. Outputs above a selected threshold $S(T, W) > 0.5$ signify that the point has been adequately corrected and can be excluded from further iterative correction. Results below the threshold designate the point for further iterative corrections. In practice we limit the maximum number of iterative corrections to a fixed constant: any corrected points still below the desired threshold at that point are accepted as-is.

4.4 Preliminary Results

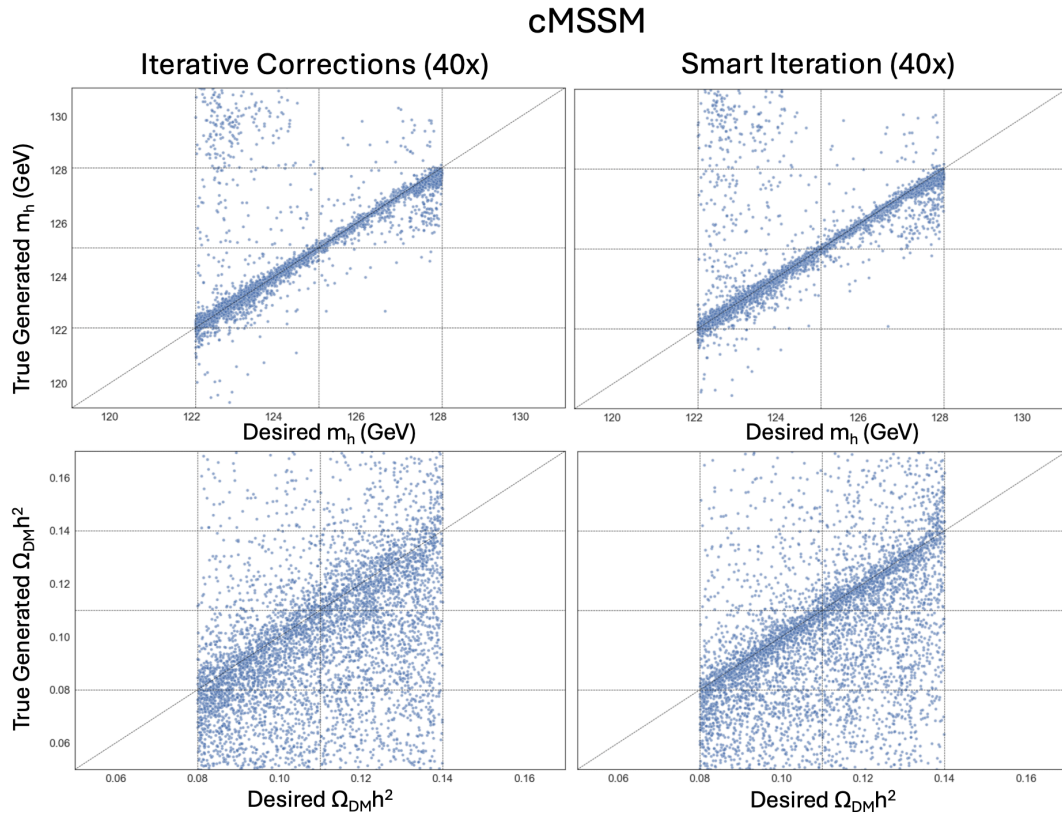


Figure 4.4: "Input/Output" plots show the results of 40x iterative corrections in the cMSSM. Standard iterative corrections are shown on the left, and 'smart iteration' on the right. The x-axis shows the input weak value W while the y-axis shows the true weak signature of the generated GUT point, calculated using SOFTSUSY and MicrOMEGAS. A 'perfect' correction would sit exactly on the diagonal $y=x$, and in practice we see a significant number of points cluster along this trend line, though this is much more prominent for the higgs constraint than for dark matter. We also note that smart iteration shows a visible improvement in dark matter performance over standard iteration, as expected.

4.4.1 cMSSM results

We observe relative success at reversing the Higgs mass constraint in the cMSSM, and while the dark matter constraint is significantly more challenging, significant inroads were made using both iterative corrections and the secondary smart iteration network, as visible in the linear clustering trends in figure 4.4. In examining the GUT-space distribution of these generated points in fig 4.5, we see the highest dark matter accuracy in regions corresponding to high density in the LOKV, such as $|A_0| \lesssim 5000$ GeV and $m_{1/2} > 1000$ GeV. In LOKV-sparse regions we note a higher density of 'theoretically invalid' points (whose GUT parameters are such that weak scale observables are not defined or calculable). Similarly, looking at weak space in fig 4.6 we note that the highest concentrations of both larger DM error and theoretical invalids occur in the LOKV-sparse vicinity $m_h \lesssim 124$ GeV (though interestingly, these seem to be fairly evenly distributed across the range of valid $(\Omega_{DM}h^2)_{Desired}$ values. While we observe a visible improvement to the DM predictability in fig 4.4 as a result of the smart iteration network, we see also that this methodology seems to have diminishing returns, as the amount of total points the network accepts begins to plateau around 41.5% after 40 iterations, as seen in fig 4.7. It is possible this could be improved with further training. In fig 4.9 we can see the presence of theoretical invalids (denoted as "-1" by our evaluation software) as anomalous peaks at $(\Omega_{DM}h^2)_{Desired} + 1$ and $(m_h)_{Desired} + 1$. Our present methodology struggles to avoid these theoretical invalids: since they have no defined weak signature and do not appear in the LOKV, the network is largely blind to their existence or position. We discuss possible remedies for this in section 5.

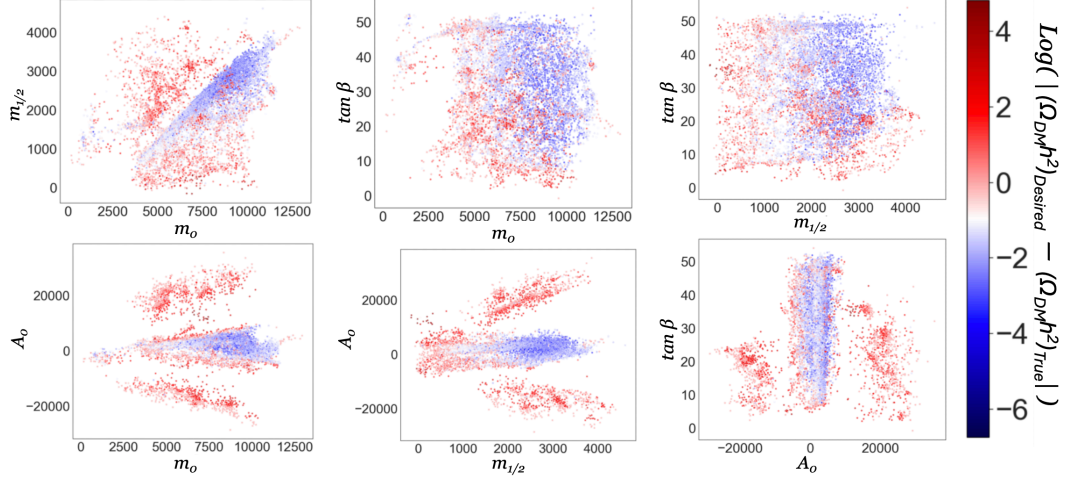


Figure 4.5: cMSSM points generated via smart iteration as seen in 4.4 are plotted in 2D slices of the cMSSM parameter space and shaded by the log of the dark matter error $\text{Log}(|(\Omega_{DM}h^2)_{Desired} - (\Omega_{DM}h^2)_{True}|)$. We observe that the model performs best in areas that correspond to high LOKV density, and worse in sparser regions.

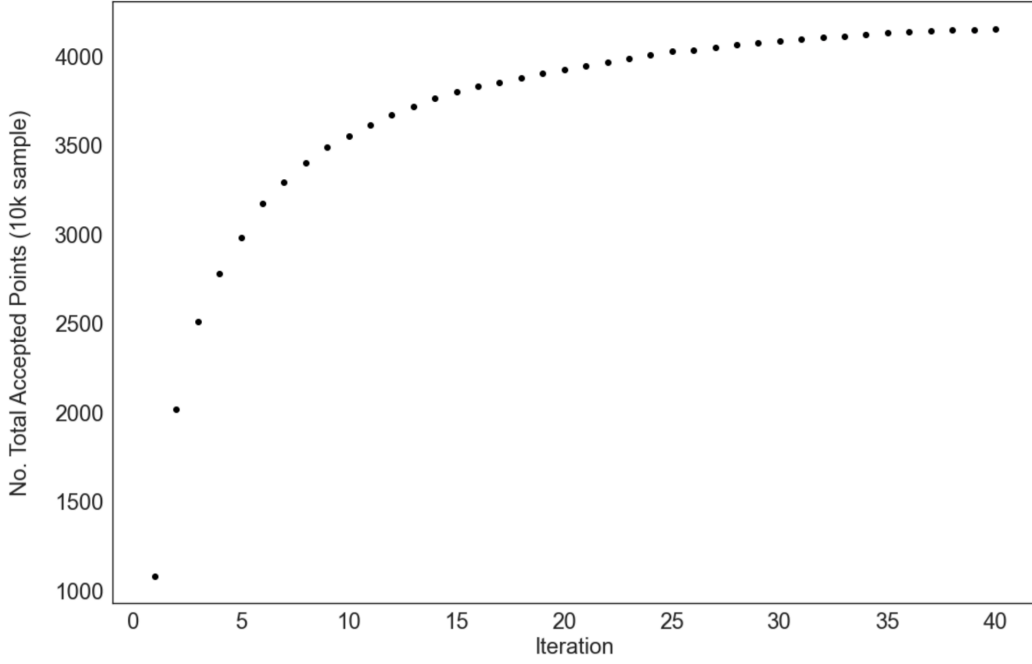


Figure 4.7: The total number of cMSSM points accepted via smart iteration is plotted as a function of the number of iterations. We see diminishing returns at around 41.5% of the 10k sample. The remaining $\mathcal{O}(5.85k)$ points are taken as-is after 40 iterations.

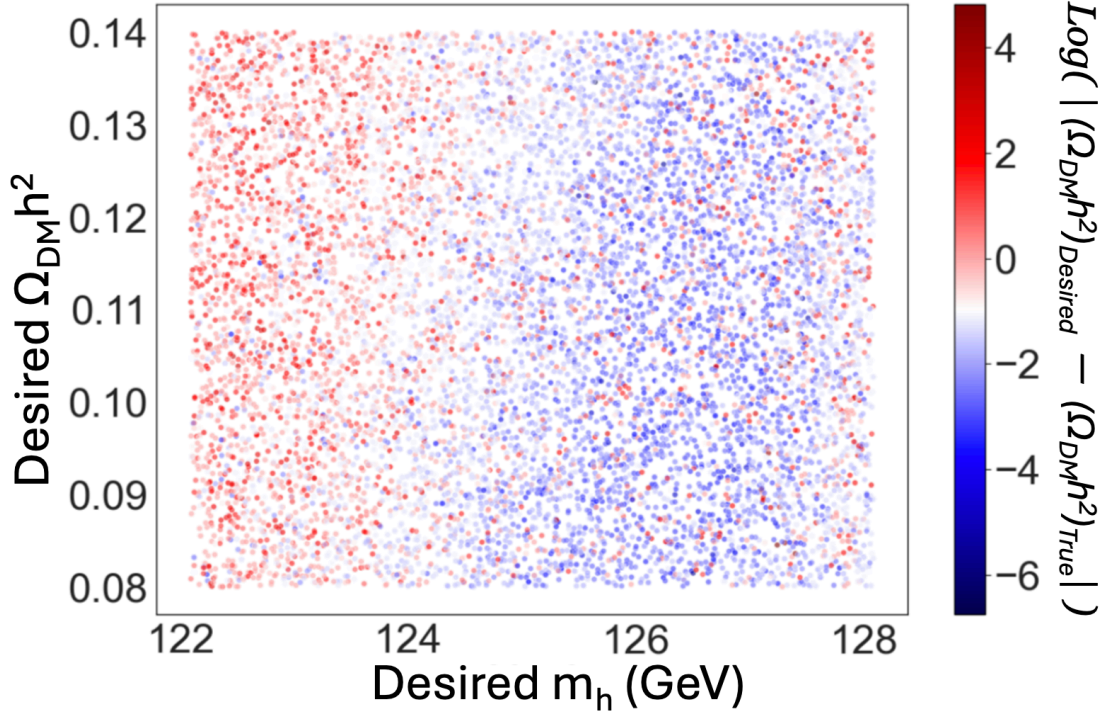


Figure 4.6: cMSSM points generated via smart iteration as seen in 4.4 are plotted in weak space and shaded by the log of the dark matter error $\text{Log}(|(\Omega_{DM}h^2)_{Desired} - (\Omega_{DM}h^2)_{True}|)$. We observe that the model performs best in areas that correspond to high LOKV density, and worse in sparser regions.

4.4.2 pMSSM results

Our results in the pMSSM are much earlier along in their development and should be viewed as strictly preliminary. Perhaps because of the increased dimensionality, the dark matter constraint seems significantly more challenging to reverse, as seen by the lack of significant clustering within the valid window in figure 4.8. However the method does perform well on the Higgs constraint even when only utilizing a single correction. Figure 4.8 shows that unlike in the cMSSM, in the pMSSM our corrections generally yield $|(\Omega_{DM}h^2)_{True}| < |(\Omega_{DM}h^2)_{Desired}|$, such that the dark matter relic density is in most cases under-saturated by the generated GUT points. We see this also in fig 4.9, where the DM error has a prominent peak around $\mathcal{O}(10^{-1})$. In comparison, in the cMSSM we see more prevalent DM over-saturation, such that there are several hundred generated points whose DM error is $\mathcal{O}(1)$ or greater. Comparing the Higgs error distributions in fig 4.9, we also see that the pMSSMrun

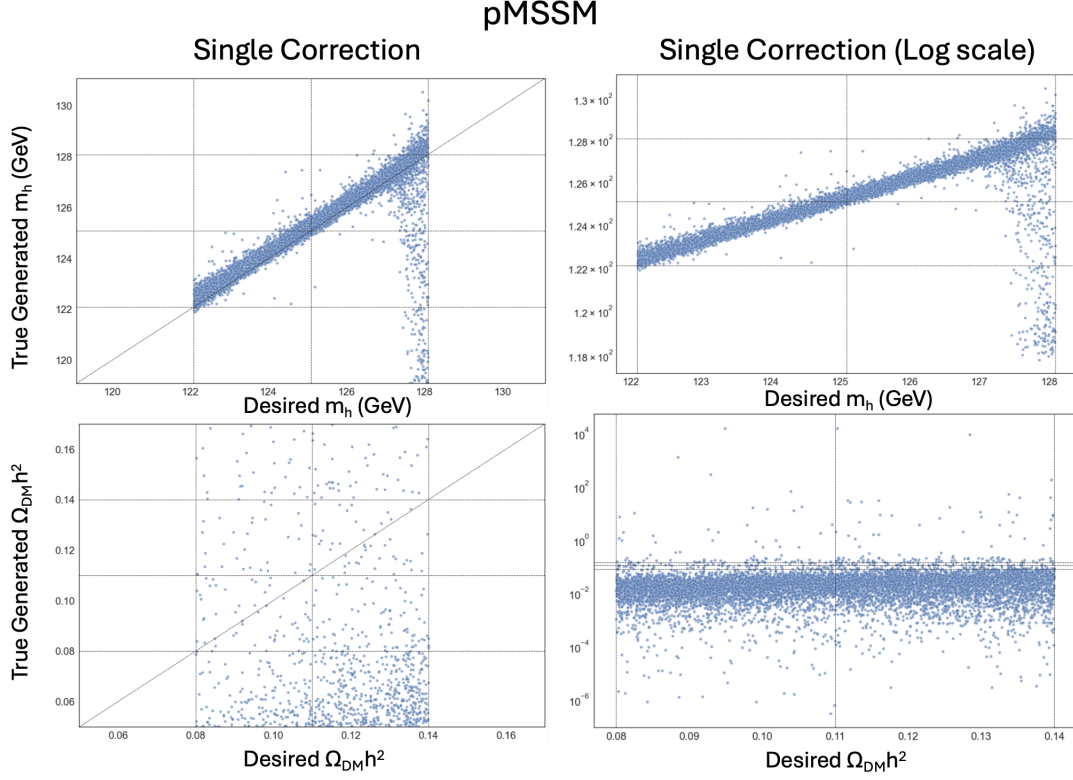


Figure 4.8: "Input/Output" plots show the results of a single correction in the pMSSM, in both standard and log scale. The x-axis shows the input weak value W while the y-axis shows the true weak signature of the generated GUT point, calculated using SOFTSUSY and MicrOMEGAS. A 'perfect' correction would sit exactly on the diagonal $y=x$, and while we see significant clustering along this trendline for the Higgs constraint, the dark matter prediction has significant inaccuracy for most generated points. Improving this is the primary remaining task for this project.

generated significantly fewer theoretical invalids than the cMSSM run. It is possible this is caused by the increase in degrees of freedom, or it may also be that theoretical invalids are most prevalent only in the valid DM window which the pMSSM model fails to accurately predict.

In general, we find sub-constraint level accuracy for m_h in both spaces, though in the pMSSM our model struggles more with the DM constraint than in the cMSSM.

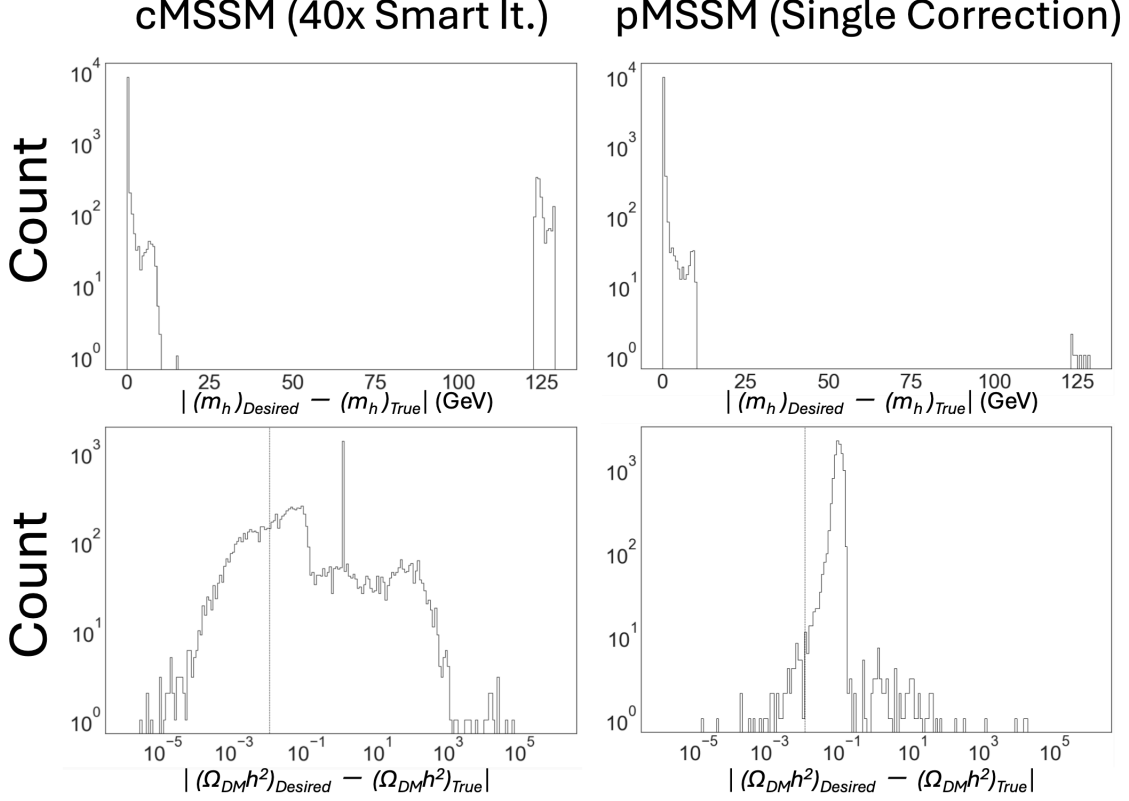


Figure 4.9: Histograms show the error across both constraints for two separate 10k runs: a 40x smart iteration run in the cMSSM on the left and a single pMSSM correction on the right. We note the following: 1.) A dotted line at $|(\Omega_{DM}h^2)_{Desired} - (\Omega_{DM}h^2)_{True}| = 0.01$ denotes our desired minimum accuracy. 2.) The dark matter predictability in the cMSSM case significantly outperforms that of the pMSSM, though both perform well on the Higgs constraint 3.) An artifact exists around $|(\Omega_{DM}h^2)_{Desired} - (\Omega_{DM}h^2)_{True}| = \mathcal{O}(1.0)$ and $|(m_h)_{Desired} - (m_h)_{True}| = \mathcal{O}(126.0)$ GeV which results from points that are 'theoretically invalid' (i.e. regions of parameter space where weak signatures are not defined/calculable). Our evaluation software marks these as "-1", creating artificial spikes corresponding to $(\Omega_{DM}h^2)_{Desired} + 1$ and $(m_h)_{Desired} + 1$. 4.) We note that the pMSSM run generated significantly fewer theoretical invalids than the cMSSM run despite using only a single correction. It is possible that theoretically invalid points are simply less common in the pMSSM due to the increase in number of degrees of freedom, or alternatively, perhaps theoretical invalids are more densely packed in regions that adequately satisfy the DM constraint.

4.5 Discussion and Conclusion

4.5.1 Future Avenues

The bulk of the remaining work in the scope of this paper consists of

1. Improving the performance of the correctional network on dark matter in the pMSSM (primarily through increasing training data size and model complexity)
2. Training a smart iteration network for the pMSSM case

In addition to this, we seek better metrics to directly quantify the accuracy of the generated point. Currently we rely primarily on visual assessments of input-output plots and error histograms, but simple averages or sums of this data fail to discriminate generally poor-performing models from those which are accurate with the exception of a few significant outlier points, especially in the case of the dark matter constraint where values (and thus errors) can sometimes reach as high as $\mathcal{O}(10^4)$. Additionally such metrics do not adequately account for the presence or absence of theoretical invalids. A suitable metric would directly enable more precise comparisons between trials.

Beyond the scope of this project, we envision that future in-roads will require adaptations which drastically reduce the presence of both theoretical invalids and volatile/outlier points. A contrastive framework and/or use of an latent embedding space [13, 121] may provide an appealing solution here (see fig 4.10). While mapping to and from such a space is an added difficulty, the potential boon of being able to perform corrections in an environment devoid of irregularities and outliers could make the core task significantly easier.

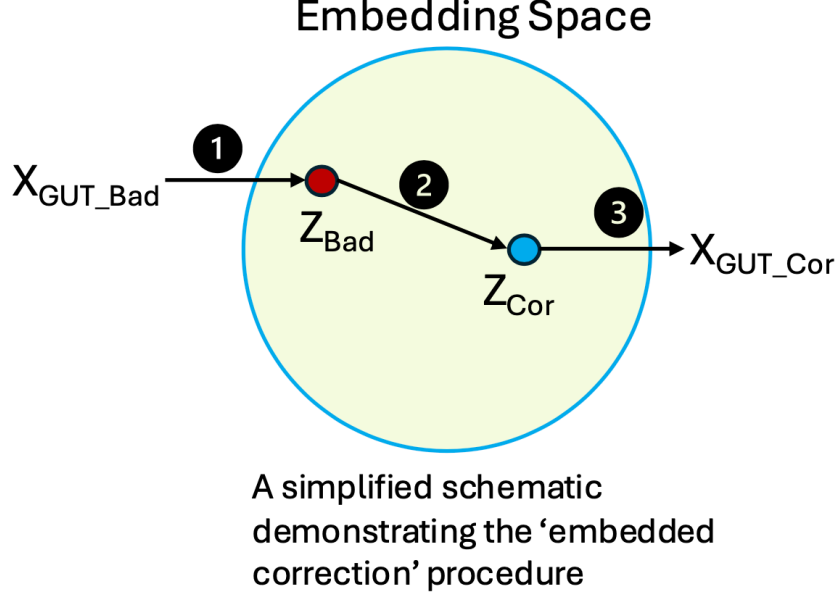


Figure 4.10: A simplified cartoon of the proposed embedding correctional network: 1.) The initial 'bad guess' points are mapped to the embedding space via an encoder. 2.) The correctional network performs the nudge strictly in latent space to yield the desired weak signature 3.) The corrected points are mapped back to GUT-space via decoder.

4.5.2 Conclusion

In conclusion, we trained a correctional network to learn how to 'nudge' a bad guess in GUT-scale parameter space such that the resulting GUT point has the desired weak signature in both the cMSSM and pMSSM. Based on preliminary results, our method performs admirably on the Higgs mass constraint in both spaces, though dark matter relic density is more challenging, especially in the pMSSM. We also trained a secondary 'smart iteration' network to monitor iterative corrections in the cMSSM with the explicit goal of improving DM accuracy. More work remains in improving the DM performance in the pMSSM. The ability to 'reverse' high-dimensional spaces this way has the potential for broader utility, not just in searching other similar spaces: if novel and interesting regions of the MSSM are discovered, it can motivate new collider analyses and particle searches, directing us to look for new physics in places we would not have otherwise.

Bibliography

- [1] G. Aad, T. Abajyan, B. Abbott, J. Abdallah, S. Abdel Khalek, A. Abdelalim, O. Abdinov, R. Aben, B. Abi, M. Abolins, and et al. Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lh. *Physics Letters B*, 716(1):1–29, 2012.
- [2] G. Aad, B. Abbott, J. Abdallah, O. Abdinov, R. Aben, M. Abolins, O. S. AbouZeid, H. Abramowicz, H. Abreu, and et al. Summary of the atlas experiment’s sensitivity to supersymmetry after lh run 1 — interpreted in the phenomenological mssm. *Journal of High Energy Physics*, 2015(10), Oct 2015.
- [3] S. Abel, D. G. Cerdeno, and S. Robles. The power of genetic algorithms: what remains of the pmssm?, 2018.
- [4] S. Abel and J. Rizos. Genetic algorithms and the search for viable string vacua. *Journal of High Energy Physics*, 2014(8), aug 2014.
- [5] B. Abi, T. Albahri, S. Al-Kilani, D. Allspach, L. P. Alonzi, A. Anastasi, A. Anisenkov, F. Azfar, K. Badgley, S. Baekler, I. Bailey, V. A. Baranov, E. Barlas-Yucel, T. Barrett, E. Barzi, A. Basti, F. Bedeschi, A. Behnke, M. Berz, M. Bhattacharya, H. P. Binney, R. Bjorkquist, P. Bloom, J. Bono, E. Bottalico, T. Bowcock, D. Boyden, G. Cantatore, R. M. Carey, J. Carroll, B. C. K. Casey, D. Cauz, S. Ceravolo, R. Chakraborty, S. P. Chang, A. Chapelain, S. Chappa, S. Charity, R. Chislett, J. Choi, Z. Chu, T. E. Chupp, M. E. Convery, A. Conway, G. Corradi, S. Corrodi, L. Crottozzi, J. D. Crnkovic, S. Dabagov, P. M. De Lurgio, P. T. Debevec, S. Di Falco, P. Di Meo, G. Di Sciascio, R. Di Stefano, B. Drendel, A. Driutti, V. N. Duginov, M. Eads, N. Eggert, A. Epps, J. Esquivel, M. Farooq, R. Fatemi, C. Ferrari, M. Fertl, A. Fiedler, A. T. Fienberg, A. Fioretti, D. Flay, S. B. Foster, H. Friedrich, E. Frlež, N. S. Froemming, J. Fry, C. Fu, C. Gabbanini, M. D. Galati, S. Ganguly, A. Garcia, D. E. Gastler, J. George, L. K. Gibbons, A. Gioiosa, K. L. Giovanetti, P. Girotti, W. Gohn, T. Gorringer, J. Grange, S. Grant, F. Gray, S. Haciomeroglu, D. Hahn, T. Halewood-Leagas, D. Hampai, F. Han, E. Hazen, J. Hempstead, S. Henry, A. T. Herrod, D. W. Hertzog, G. Hesketh, A. Hibbert, Z. Hodge, J. L. Holzbauer, K. W. Hong, R. Hong, M. Iacovacci, M. Incagli, C. Johnstone, J. A. Johnstone, P. Kammel, M. Kargiantoulakis, M. Karuza, J. Kaspar, D. Kawall, L. Kelton, A. Keshavarzi, D. Kessler, K. S. Khaw, Z. Khechadorian, N. V. Khomutov, B. Kiburg, M. Kiburg, O. Kim, S. C. Kim, Y. I. Kim, B. King,

- N. Kinnaird, M. Korostelev, I. Kourbanis, E. Kraegelloh, V. A. Krylov, A. Kuchibhotla, N. A. Kuchinskiy, K. R. Labe, J. LaBounty, M. Lancaster, M. J. Lee, S. Lee, S. Leo, B. Li, D. Li, L. Li, I. Logashenko, A. Lorente Campos, A. Lucà, G. Lukicov, G. Luo, A. Lusiani, A. L. Lyon, B. MacCoy, R. Madrak, K. Makino, F. Marignetti, S. Mastroianni, S. Maxfield, M. McEvoy, W. Merritt, A. A. Mikhailichenko, J. P. Miller, S. Miozzi, J. P. Morgan, W. M. Morse, J. Mott, E. Motuk, A. Nath, D. Newton, H. Nguyen, M. Oberling, R. Osofsky, J.-F. Ostiguy, S. Park, G. Pauletta, G. M. Piacentino, R. N. Pilato, K. T. Pitts, B. Plaster, D. Počanić, N. Pohlman, C. C. Polly, M. Popovic, J. Price, B. Quinn, N. Raha, S. Ramachandran, E. Ramberg, N. T. Rider, J. L. Ritchie, B. L. Roberts, D. L. Rubin, L. Santi, D. Sathyan, H. Schellman, C. Schlesier, A. Schreckenberger, Y. K. Semertzidis, Y. M. Shatunov, D. Shemyakin, M. Shenk, D. Sim, M. W. Smith, A. Smith, A. K. Soha, M. Sorbara, D. Stöckinger, J. Stapleton, D. Still, C. Stoughton, D. Stratakis, C. Strohman, T. Stuttard, H. E. Swanson, G. Sweetmore, D. A. Sweigart, M. J. Syphers, D. A. Tarazona, T. Teubner, A. E. Tewsley-Booth, K. Thomson, V. Tishchenko, N. H. Tran, W. Turner, E. Valtov, D. Vasilkova, G. Venanzoni, V. P. Volnykh, T. Walton, M. Warren, A. Weiskopf, L. Welty-Rieger, M. Whitley, P. Winter, A. Wolski, M. Wormald, W. Wu, and C. Yoshikawa. Measurement of the positive muon anomalous magnetic moment to 0.46 ppm. *Phys. Rev. Lett.*, 126:141801, Apr 2021.
- [6] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski. A learning algorithm for boltzmann machines. *Cognitive Science*, 9(1):147–169, 1985.
- [7] A. G. Akeroyd and F. Mahmoudi. Measuring vuband probing SUSY with double ratios of purely leptonic decays of B and D mesons. *Journal of High Energy Physics*, 2010(10):38, Oct. 2010.
- [8] B. Allanach. Softsusy: A program for calculating supersymmetric spectra. *Computer Physics Communications*, 143(3):305–331, 2002.
- [9] B. C. Allanach. Impact of cms multi-jets and missing energy search on cmssm fits. *Phys. Rev. D*, 83:095019, May 2011.
- [10] C. Alós-Ferrer and M. Mihm. An axiomatic characterization of bayesian updating. *Journal of Mathematical Economics*, 104:102799, 12 2022.
- [11] E. A. Baltz and P. Gondolo. Markov chain monte carlo exploration of minimal supergravity with implications for dark matter. *Journal of High Energy Physics*, 2004(10):052–052, Oct 2004.
- [12] R. Barbieri and A. Strumia. The ‘lep paradox’, 2000.
- [13] J. Baretz, N. Carrara, J. Hollingsworth, and D. Whiteson. Visualization and efficient generation of constrained high-dimensional theoretical parameter spaces. *Journal of High Energy Physics*, 2023(11):62, Nov. 2023.
- [14] J. B. Baretz, M. Fieg, V. Ganesh, A. Ghosh, V. Knapp-Perez, J. Rudolph, and D. Whiteson. Towards ai-assisted neutrino flavor theory design, 2025.

- [15] P. Bechtle, S. Heinemeyer, O. Stål, T. Stefaniak, and G. Weiglein. Higgs signals: Confronting arbitrary higgs sectors with measurements at the tevatron and the lhc. *The European Physical Journal C*, 74(2), Feb. 2014.
- [16] G. Belanger, F. Boudjema, A. Pukhov, and A. Semenov. micromegas: A tool for dark matter studies. *Il Nuovo Cimento C*, 33(2):111–116, Aug 2010.
- [17] 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, and E. L. Wright. NINE-YEAR WILKINSON MICROWAVE ANISOTROPY PROBE (WMAP) OBSERVATIONS: FINAL MAPS AND RESULTS. *The Astrophysical Journal Supplement Series*, 208(2):20, sep 2013.
- [18] G. W. Bennett, B. Bousquet, H. N. Brown, G. Bunce, R. M. Carey, P. Cushman, G. T. Danby, P. T. Debevec, M. Deile, H. Deng, W. Deninger, S. K. Dhawan, V. P. Druzhinin, L. Duong, E. Efstathiadis, F. J. M. Farley, G. V. Fedotovitch, S. Giron, F. E. Gray, D. Grigoriev, M. Grosse-Perdekamp, A. Grossmann, M. F. Hare, D. W. Hertzog, X. Huang, V. W. Hughes, M. Iwasaki, K. Jungmann, D. Kawall, M. Kawamura, B. I. Khazin, J. Kindem, F. Krienen, I. Kronkvist, A. Lam, R. Larsen, Y. Y. Lee, I. Logashenko, R. McNabb, W. Meng, J. Mi, J. P. Miller, Y. Mizumachi, W. M. Morse, D. Nikas, C. J. G. Onderwater, Y. Orlov, C. S. Özben, J. M. Paley, Q. Peng, C. C. Polly, J. Pretz, R. Prigl, G. zu Putlitz, T. Qian, S. I. Redin, O. Rind, B. L. Roberts, N. Ryskulov, S. Sedykh, Y. K. Semertzidis, P. Shagin, Y. M. Shatunov, E. P. Sichtermann, E. Solodov, M. Sossong, A. Steinmetz, L. R. Sulak, C. Timmermans, A. Trofimov, D. Urner, P. von Walter, D. Warburton, D. Winn, A. Yamamoto, and D. Zimmerman. Final report of the e821 muon anomalous magnetic moment measurement at bnl. *Phys. Rev. D*, 73:072003, Apr 2006.
- [19] M. Betancourt. A conceptual introduction to hamiltonian monte carlo, 2018.
- [20] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2006.
- [21] P. G. Bissiri, C. C. Holmes, and S. G. Walker. A general framework for updating belief distributions. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 78(5):1103–1130, Feb. 2016.
- [22] J. Brehmer and K. Cranmer. Flows for simultaneous manifold learning and density estimation, 2020.
- [23] M. Bridges, K. Cranmer, F. Feroz, M. Hobson, R. Ruiz de Austri, and R. Trotta. A coverage study of the cmssm based on atlas sensitivity using fast neural networks techniques. *Journal of High Energy Physics*, 2011(3), Mar 2011.
- [24] M. Bridges, K. Cranmer, and F. Feroz et al. A coverage study of the CMSSM based on ATLAS sensitivity using fast neural networks techniques. *J. High Energ. Phys.*, (12), 2011.

- [25] O. Buchmueller, R. Cavanaugh, A. Roeck, and et al. The CMSSM and NUHM1 after LHC Run 1. *Eur. Phys. J. C*, 74(2922), 2014.
- [26] M. R. Buckley, D. Feld, S. Macaluso, A. Monteux, and D. Shih. Cornering natural SUSY at LHC run II and beyond. *Journal of High Energy Physics*, 2017(8), aug 2017.
- [27] M. R. Buckley, A. Monteux, and D. Shih. Precision corrections to fine tuning in SUSY. *Journal of High Energy Physics*, 2017(6), jun 2017.
- [28] M. Cahill-Rowley, J. L. Hewett, A. Ismail, and T. G. Rizzo. pmssm studies at the 7, 8 and 14 tev lhc, 2013.
- [29] M. Cahill-Rowley, J. L. Hewett, A. Ismail, and T. G. Rizzo. pmssm studies at the 7, 8 and 14 tev lhc, 2013.
- [30] M. Cahill-Rowley, J. L. Hewett, A. Ismail, and T. G. Rizzo. Lessons and prospects from the pmssm after lhc run i. *Phys. Rev. D*, 91:055002, Mar 2015.
- [31] S. Caron, T. Heskes, S. Otten, and B. Stienen. Constraining the parameters of high-dimensional models with active learning. *The European Physical Journal C*, 79(11), Nov. 2019.
- [32] S. Caron, J. S. Kim, K. Rolbiecki, R. R. de Austri, and B. Stienen. The bsm-ai project: Susy-ai-generalizing lhc limits on supersymmetry with machine learning. *The European Physical Journal C*, 77(4):257, Apr 2017.
- [33] N. Carrara. Constraint-driven high-dimensional uncompressed (categorical) clustering (chuncnet). <https://github.com/infophysics/CHUNCNet>, 2023.
- [34] F. Carta, A. Gauntlett, F. Griffin, and Y.-H. He. Bps spectroscopy with reinforcement learning, 2025.
- [35] D. Cerd  o, M. Peir  , and S. Robles. Fits to the fermi-lat gev excess with right-handed sneutrino dark matter: Implications for direct and indirect dark matter searches and the lhc. *Physical Review D*, 91(12), 2015.
- [36] A. H. Chamseddine, R. Arnowitt, and P. Nath. Locally supersymmetric grand unification. *Phys. Rev. Lett.*, 49:970–974, Oct 1982.
- [37] S. Chatrchyan, V. Khachatryan, A. Sirunyan, A. Tumasyan, W. Adam, E. Aguilo, T. Bergauer, M. Dragicevic, J. Er  , C. Fabjan, and et al. Study of the mass and spin-parity of the higgs boson candidate via its decays to z boson pairs. *Physical Review Letters*, 110(8), Feb. 2013.
- [38] A. Chatterjee, A. Choudhury, S. Mitra, A. Mondal, and S. Mondal. Exploring the bsm parameter space with neural network aided simulation-based inference, 2025.
- [39] A. Chatterjee, A. Choudhury, S. Mitra, A. Mondal, and S. Mondal. Exploring the bsm parameter space with neural network aided simulation-based inference. *Journal of High Energy Physics*, 2025(12), Dec. 2025.

- [40] C. Cheung, L. J. Hall, D. Pinner, and J. T. Ruderman. Prospects and Blind Spots for Neutralino Dark Matter. *JHEP*, 05:100, 2013.
- [41] T. Cohen and J. G. Wacker. Here be dragons: the unexplored continents of the CMSSM. *Journal of High Energy Physics*, 2013(9), sep 2013.
- [42] T. C. collaboration., V. Khachatryan, A. Sirunyan, and et al. Phenomenological MSSM interpretation of CMS searches in pp collisions at $\sqrt{s} = 7$ and 8 TeV. *J. High Energ. Phys.*, (129), 2016.
- [43] T. M. Cover and J. A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, USA, 2006.
- [44] F. A. de Souza, M. C. Romão, N. F. Castro, M. Nikjoo, and W. Porod. Exploring parameter spaces with artificial intelligence and machine learning black-box optimisation algorithms, 2022.
- [45] M. Di Marzio and G. Lafratta. Reducing dimensionality effects on kernel density estimation: The bivariate gaussian case. In M. Vichi and O. Opitz, editors, *Classification and Data Analysis*, pages 287–294, Berlin, Heidelberg, 1999. Springer Berlin Heidelberg.
- [46] J. Dickinson, S. Bein, S. Heinemeyer, J. Hiltbrand, J. Hirschauer, W. Hopkins, E. Lipeles, M. Mrowietz, and N. Strobbe. A grand scan of the pmssm parameter space for snowmass 2021, 2022.
- [47] M. Dine and A. E. Nelson. Dynamical supersymmetry breaking at low energies. *Physical Review D*, 48(3):1277–1287, aug 1993.
- [48] M. Dine, A. E. Nelson, Y. Nir, and Y. Shirman. New tools for low energy dynamical supersymmetry breaking. *Physical Review D*, 53(5):2658–2669, mar 1996.
- [49] M. Dine, A. E. Nelson, and Y. Shirman. Low energy dynamical supersymmetry breaking simplified. *Physical Review D*, 51(3):1362–1370, feb 1995.
- [50] A. Djouadi et al. The Minimal supersymmetric standard model: Group summary report. In *GDR (Groupement De Recherche) - Supersymetrie*, 12 1998.
- [51] A. Djouadi, S. Rosier-Lees, M. Bezouh, M. A. Bizouard, C. Boehm, F. Borzumati, C. Briot, J. Carr, M. B. Causse, F. Charles, X. Chereau, P. Colas, L. DufLOT, A. Dupperin, A. Ealet, H. El-Mamouni, N. Ghodbane, F. Gieres, B. Gonzalez-Pineiro, S. Gourmelen, G. Grenier, P. Gris, J. F. Grivaz, C. Hebrard, B. Ille, J. L. Kneur, N. Kostantinidis, J. Layssac, P. Lebrun, R. Ledu, M. C. Lemaire, C. LeMouel, L. Luginier, Y. Mambrini, J. P. Martin, G. Montarou, G. MoultaKa, S. Muanza, E. Nuss, E. Perez, F. M. Renard, D. Reynaud, L. Serin, C. Thevenet, A. Trabelsi, F. Zach, and D. Zerwas. The minimal supersymmetric standard model: Group summary report, 1999.
- [52] C. Durkan, A. Bekasov, I. Murray, and G. Papamakarios. Neural spline flows. 2019.

- [53] J. Ellis, S. King, and J. Roberts. The fine-tuning price of neutralino dark matter in models with non-universal higgs masses. *Journal of High Energy Physics*, 2008(04):099–099, Apr. 2008.
- [54] J. A. Evans, Y. Kats, D. Shih, and M. J. Strassler. Toward full LHC coverage of natural supersymmetry. *Journal of High Energy Physics*, 2014(7), jul 2014.
- [55] W. J. FAWCETT. pMSSM studies with ATLAS and CMS. In *Proceedings of Fourth Annual Large Hadron Collider Physics — PoS(LHCP2016)*, volume 276, page 146, 2016.
- [56] C. Fefferman, S. Mitter, and H. Narayanan. Testing the manifold hypothesis. *Journal of the American Mathematical Society*, 29(4):983–1049, Oct 2016.
- [57] J. L. Feng, K. T. Matchev, and T. Moroi. Focus points and naturalness in supersymmetry. *Phys. Rev. D*, 61:075005, 2000.
- [58] J. L. Feng and D. Sanford. Natural 125 gev higgs boson in the mssm from focus point supersymmetry with a -terms. *Phys. Rev. D*, 86:055015, Sep 2012.
- [59] F. Feruglio. Fermion masses, critical behavior and universality, 2023.
- [60] K. Freese, A. López, N. R. Shah, and B. Shakya. Mssm a -funnel and the galactic center excess: prospects for the lhc and direct detection experiments. *Journal of High Energy Physics*, 2016(4):1–28, Apr. 2016.
- [61] E. Fuentes. Sliced-wasserstein autoencoder - pytorch. <https://github.com/eifuentes/swae-pytorch>, 2018.
- [62] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, 0 edition, Nov. 2013.
- [63] D. Ghosh, M. Guchait, S. Raychaudhuri, and D. Sengupta. How constrained is the constrained mssm? *Phys. Rev. D*, 86:055007, Sep 2012.
- [64] D. Ghosh, M. Guchait, S. Raychaudhuri, and D. Sengupta. How constrained is the constrained mssm? *Physical Review D*, 86(5), Sep 2012.
- [65] C. R. Givens and R. M. Shortt. A class of wasserstein metrics for probability distributions. *Michigan Mathematical Journal*, 31:231–240, 1984.
- [66] D. E. Goldberg. Genetic algorithms in search optimization and machine learning. 1988.
- [67] T. Golling, S. Klein, R. Mastandrea, B. Nachman, and J. A. Raine. Morphing one dataset into another with maximum likelihood estimation. *Physical Review D*, 108(9), Nov. 2023.
- [68] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial networks, 2014.

- [69] P. W. Graham, H. Ramani, and S. S. Y. Wong. Enhancing direct detection of higgsino dark matter, 2026.
- [70] H. F. A. Group, Y. Amhis, S. Banerjee, E. Ben-Haim, S. Blyth, A. Bozek, C. Bozzi, A. Carbone, R. Chistov, M. Chrzyszcz, G. Cibinetto, J. Dingfelder, M. Gelb, M. Gersabeck, T. Gershon, L. Gibbons, B. Golob, R. Harr, K. Hayasaka, H. Hayashii, T. Kuhr, O. Leroy, A. Lusiani, K. Miyabayashi, P. Naik, S. Nishida, A. O. Campos, M. Patel, D. Pedrini, M. Petrič, M. Rama, M. Roney, M. Rotondo, O. Schneider, C. Schwanda, A. J. Schwartz, B. Shwartz, J. G. Smith, R. Tesarek, K. Trabelsi, P. Urquijo, R. V. Kooten, and A. Zupanc. Averages of b -hadron, c -hadron, and τ -lepton properties as of summer 2014, 2014.
- [71] L. J. Hall, D. Pinner, and J. T. Ruderman. A natural susy higgs near 125 gev. *Journal of High Energy Physics*, 2012:1–25, 2011.
- [72] J. Halverson, B. Nelson, and F. Ruehle. Branes with brains: exploring string vacua with deep reinforcement learning. *Journal of High Energy Physics*, 2019(6), 2019.
- [73] C. Han, K. ichi Hikasa, L. Wu, J. M. Yang, and Y. Zhang. Status of cmssm in light of current lhc run-2 and lux data. *Physics Letters B*, 769:470–476, 2017.
- [74] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. Fernández del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant. Array programming with NumPy. *Nature*, 585:357–362, 2020.
- [75] T. R. Harvey and A. Lukas. Particle physics model building with reinforcement learning, 2021.
- [76] Y.-H. He and J. M. P. Ipiña. Machine-learning the classification of spacetimes, 2022.
- [77] J. Hollingsworth, M. Ratz, P. Tanedo, and D. Whiteson. Efficient sampling of constrained high-dimensional theoretical spaces with machine learning. *The European Physical Journal C*, 81(12), dec 2021.
- [78] J. N. Howard, S. Mandt, D. Whiteson, and Y. Yang. Learning to simulate high energy particle collisions from unlabeled data. *Scientific Reports*, 12(1), may 2022.
- [79] P. Huang and C. E. M. Wagner. Blind spots for neutralino dark matter in the mssm with an intermediate $m_{\tilde{a}}$. *Physical Review D*, 90:015018, 2014.
- [80] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in science & engineering*, 9(3):90–95, 2007.
- [81] H. Itoh, S. Komine, and Y. Okada. Tauonic b decays in the minimal supersymmetric standard model. *Progress of Theoretical Physics*, 114(1):179–204, 2005.

- [82] V. Khachatryan, A. M. Sirunyan, A. Tumasyan, W. Adam, E. Asilar, T. Bergauer, J. Brandstetter, E. Brondolin, M. Dragicevic, and et al. Phenomenological mssm interpretation of cms searches in pp collisions at $\sqrt{s} = 7$ and 8 tev. *Journal of High Energy Physics*, 2016(10), Oct 2016.
- [83] V. Khachatryan, A. M. Sirunyan, A. Tumasyan, W. Adam, T. Bergauer, M. Dragicevic, J. Erö, M. Friedl, R. Frühwirth, V. M. Ghete, C. Hartl, N. Hörmann, J. Hrubec, M. Jeitler, W. Kiesenhofer, V. Knünz, M. Krammer, I. Krätschmer, D. Liko, I. Mikulec, D. Rabady, B. Rahbaran, H. Rohringer, R. Schöfbeck, J. Strauss, W. Treberer-Treberspur, W. Waltenberger, C.-E. Wulz, V. Mossolov, N. Shumeiko, J. Suarez Gonzalez, S. Alderweireldt, S. Bansal, T. Cornelis, E. A. De Wolf, X. Janssen, A. Knutsson, J. Lauwers, S. Luyckx, S. Ochesanu, R. Rougny, M. Van De Klundert, H. Van Haevermaet, P. Van Mechelen, N. Van Remortel, A. Van Spilbeeck, F. Blekman, S. Blyweert, J. D'Hondt, N. Daci, N. Heracleous, J. Keaveney, S. Lowette, M. Maes, A. Olbrechts, Q. Python, D. Strom, S. Tavernier, W. Van Doninck, P. Van Mulders, G. P. Van Onsem, I. Villella, C. Caillol, B. Clerbaux, G. De Lentdecker, D. Dobur, L. Favart, A. P. R. Gay, A. Grebenyuk, A. Léonard, A. Mohammadi, L. Perniè, A. Randleconde, T. Reis, T. Seva, L. Thomas, C. Vander Velde, P. Vanlaer, J. Wang, F. Zenoni, V. Adler, K. Beernaert, L. Benucci, A. Cimmino, S. Costantini, S. Crucy, S. Dildick, A. Fagot, G. Garcia, J. Mccartin, A. A. Ocampo Rios, D. Ryckbosch, S. Salva Diben, M. Sigamani, N. Strobbe, F. Thyssen, M. Tytgat, E. Yazgan, N. Zaganidis, S. Basegmez, C. Beluffi, G. Bruno, R. Castello, A. Caudron, L. Ceard, G. G. Da Silveira, C. Delaere, T. du Pree, D. Favart, L. Forthomme, A. Giammanco, J. Hollar, A. Jafari, P. Jez, M. Komm, V. Lemaitre, C. Nuttens, D. Pagano, L. Perrini, A. Pin, K. Piotrkowski, A. Popov, L. Quertenmont, M. Selvaggi, M. Vidal Marono, J. M. Vizan Garcia, N. Beliy, T. Caebergs, E. Daubie, G. H. Hammad, W. L. Aldá Júnior, G. A. Alves, L. Brito, M. Correa Martins Junior, T. Dos Reis Martins, C. Mora Herrera, M. E. Pol, P. Rebello Teles, W. Carvalho, J. Chinellato, A. Custódio, E. M. Da Costa, D. De Jesus Damiao, C. De Oliveira Martins, S. Fonseca De Souza, H. Malbouisson, D. Matos Figueiredo, L. Mundim, H. Nogima, W. L. Prado Da Silva, J. Santaolalla, A. Santoro, A. Sznajder, E. J. Tonelli Manganote, A. Vilela Pereira, C. A. Bernardes, S. Dogra, T. R. Fernandez Perez Tomei, E. M. Gregores, P. G. Mercadante, S. F. Novaes, S. S. Padula, A. Aleksandrov, V. Genchev, R. Hadjiiska, P. Iaydjiev, A. Marinov, S. Piperov, M. Rodozov, G. Sultanov, M. Vutova, A. Dimitrov, I. Glushkov, L. Litov, B. Pavlov, P. Petkov, J. G. Bian, G. M. Chen, H. S. Chen, M. Chen, T. Cheng, R. Du, C. H. Jiang, R. Plestina, F. Romeo, J. Tao, Z. Wang, C. Asawatangtrakuldee, Y. Ban, Q. Li, S. Liu, Y. Mao, S. J. Qian, D. Wang, Z. Xu, W. Zou, C. Avila, A. Cabrera, L. F. Chaparro Sierra, C. Florez, J. P. Gomez, B. Gomez Moreno, J. C. Sanabria, N. Godinovic, D. Lelas, D. Polic, I. Puljak, Z. Antunovic, M. Kovac, V. Brigljevic, K. Kadija, J. Luetic, D. Mekterovic, L. Sudic, A. Attikis, G. Mavromanolakis, J. Mousa, C. Nicolaou, F. Ptochos, P. A. Razis, M. Bodlak, M. Finger, M. Finger, Jr, Y. Assran, A. Ellithi Kamel, M. A. Mahmoud, A. Radi, M. Kadastik, M. Murumaa, M. Raidal, A. Tiko, P. Eerola, G. Fedi, M. Voutilainen, J. Härkönen, V. Karimäki, R. Kinnunen, M. J. Kortelainen, T. Lampén, K. Lassila-Perini, S. Lehti, T. Lindén, P. Luukka, T. Mäenpää, T. Peltola, E. Tuominen, J. Tuominiemi, E. Tuovinen, L. Wendland,

- J. Talvitie, T. Tuuva, M. Besancon, F. Couderc, M. Dejardin, D. Denegri, B. Fabbro, J. L. Faure, C. Favaro, F. Ferri, S. Ganjour, A. Givernaud, P. Gras, G. Hamel de Monchenault, P. Jarry, E. Locci, J. Malcles, J. Rander, A. Rosowsky, M. Titov, S. Baffioni, F. Beaudette, P. Busson, C. Charlot, T. Dahms, M. Dalchenko, L. Dobrzynski, N. Filipovic, A. Florent, R. Granier de Cassagnac, L. Mastrolorenzo, P. Miné, C. Mironov, I. N. Naranjo, M. Nguyen, C. Ochando, G. Ortona, P. Paganini, S. Regnard, R. Salerno, J. B. Sauvan, Y. Sirois, C. Veelken, Y. Yilmaz, A. Zabi, J.-L. Agram, J. Andrea, A. Aubin, D. Bloch, J.-M. Brom, and CMS Collaboration. Observation of the rare $bs0 \rightarrow \mu + \mu^-$ decay from the combined analysis of CMS and LHCb data. *Nature*, 522(7554):68–72, June 2015.
- [84] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization, 2014.
- [85] D. P. Kingma and M. Welling. Auto-encoding variational bayes, 2013.
- [86] D. Kirkby. Machine learning statistics. <https://github.com/dkirkby/MachineLearningStatistics>, 2021.
- [87] S. Klein, J. A. Raine, and T. Golling. Flows for flows: Training normalizing flows between arbitrary distributions with maximum likelihood estimation, 2022.
- [88] I. Kobyzev, S. J. Prince, and M. A. Brubaker. Normalizing flows: An introduction and review of current methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(11):3964–3979, nov 2021.
- [89] T. Kohonen. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43:59–69, 2004.
- [90] T. Kohonen and T. Honkela. Kohonen network. *Scholarpedia*, 2(1):1568, Jan. 2007.
- [91] S. Kolouri, S. R. Park, M. Thorpe, D. Slepcev, and G. K. Rohde. Optimal mass transport: Signal processing and machine-learning applications. *IEEE Signal Processing Magazine*, 34(4):43–59, 2017.
- [92] S. Kolouri, P. E. Pope, C. E. Martin, and G. K. Rohde. Sliced-wasserstein autoencoder: An embarrassingly simple generative model, 2018.
- [93] Y. Kossale, M. Airaj, and A. Darouichi. Mode collapse in generative adversarial networks: An overview. In *2022 8th International Conference on Optimization and Applications (ICOA)*, pages 1–6, 2022.
- [94] K. Kowalska, L. Roszkowski, E. M. Sessolo, and S. Trojanowski. Low fine tuning in the mssm with higgsino dark matter and unification constraints. *Journal of High Energy Physics*, 2014:1–36, 2014.
- [95] K. Kowalska and E. M. Sessolo. Natural mssm after the lhc 8 tev run. *Physical Review D*, 88:075001, 2013.

- [96] B. Kronheim, M. Kuchera, H. Prosper, and A. Karbo. Bayesian neural networks for fast susy predictions. *Physics Letters B*, 813:136041, Feb 2021.
- [97] A. Kumar and M. Sarovar. Manifold learning via quantum dynamics, 2022.
- [98] A. Kumar and M. Sarovar. Shining light on data: Geometric data analysis through quantum dynamics, 2022.
- [99] Y. Lipman, R. T. Q. Chen, H. Ben-Hamu, M. Nickel, and M. Le. Flow matching for generative modeling. Oct. 2022.
- [100] S. P. MARTIN. A SUPERSYMMETRY PRIMER. In *Perspectives on Supersymmetry*, pages 1–98. WORLD SCIENTIFIC, jul 1998.
- [101] L. McInnes, J. Healy, and J. Melville. Umap: Uniform manifold approximation and projection for dimension reduction, 2018.
- [102] W. McKinney et al. Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, volume 445, pages 51–56. Austin, TX, 2010.
- [103] P. Meade, N. Seiberg, and D. Shih. General gauge mediation. *Progress of Theoretical Physics Supplement*, 177:143–158, 2009.
- [104] L. Morrison, S. Profumo, and J. Tamanas. Simulation based inference for efficient theory space sampling: an application to supersymmetric explanations of the anomalous muon ($g-2$), 2022.
- [105] A. Mütter, E. Parr, and P. K. Vaudrevange. Deep learning in the heterotic orbifold landscape. *Nuclear Physics B*, 940:113–129, mar 2019.
- [106] T. Nagler and C. Czado. Evading the curse of dimensionality in nonparametric density estimation with simplified vine copulas. *Journal of Multivariate Analysis*, 151:69–89, oct 2016.
- [107] R. M. Neal. Mcmc using hamiltonian dynamics, 2012.
- [108] S. Nishimura, C. Miyao, and H. Otsuka. Exploring the flavor structure of quarks and leptons with reinforcement learning. *Journal of High Energy Physics*, 2023(12), Dec. 2023.
- [109] S. Nishimura, C. Miyao, and H. Otsuka. Reinforcement learning-based statistical search strategy for an axion model from flavor, 2025.
- [110] G. Papamakarios, T. Pavlakou, and I. Murray. Masked autoregressive flow for density estimation. 2018.
- [111] M. Papucci, J. T. Ruderman, and A. Weiler. Natural SUSY Endures. *JHEP*, 09:035, 2012.

- [112] S. E. Park, P. Harris, and B. Ostdiek. Neural embedding: Learning the embedding of the manifold of physics data, 2022.
- [113] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.
- [114] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [115] Y. Ren, H. Zhao, Y. Khoo, and L. Ying. High-dimensional density estimation with tensorizing flow, 2022.
- [116] H. E. Robbins and S. Monro. A stochastic approximation method. *Annals of Mathematical Statistics*, 22:400–407, 1951.
- [117] Y. Rubner, C. Tomasi, and L. Guibas. A metric for distributions with applications to image databases. In *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, pages 59–66, 1998.
- [118] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, Oct. 1986.
- [119] F. Schroff, D. Kalenichenko, and J. Philbin. FaceNet: A unified embedding for face recognition and clustering. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2015.
- [120] M. Schultz and T. Joachims. Learning a distance metric from relative comparisons. In S. Thrun, L. Saul, and B. Schölkopf, editors, *Advances in Neural Information Processing Systems*, volume 16. MIT Press, 2003.
- [121] A. Shmakov, K. T. Greif, M. J. Fenton, A. Ghosh, P. Baldi, and D. Whiteson. Full event particle-level unfolding with variable-length latent variational diffusion. *SciPost Physics*, 18(4), Apr. 2025.
- [122] D. N. Spergel, L. Verde, H. V. Peiris, E. Komatsu, M. R. Nolta, C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, and et al. First-year wilkinson microwave anisotropy probe (wmap) observations: Determination of cosmological parameters. *The Astrophysical Journal Supplement Series*, 148(1):175–194, Sep 2003.
- [123] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

- [124] N. Tishby and N. Zaslavsky. Deep learning and the information bottleneck principle, 2015.
- [125] I. Tolstikhin, O. Bousquet, S. Gelly, and B. Schoelkopf. Wasserstein auto-encoders, 2019.
- [126] L. van der Maaten and G. Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(86):2579–2605, 2008.
- [127] G. Van Rossum and F. L. Drake. *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA, 2009.
- [128] M. L. Waskom. seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60):3021, 2021.
- [129] N. Whiteley, A. Gray, and P. Rubin-Delanchy. Statistical exploration of the manifold hypothesis, 2025.
- [130] G. N. Wojcik, S. T. Eu, and L. L. Everett. Graph reinforcement learning for exploring model spaces beyond the standard model. *Physical Review D*, 111(3), Feb. 2025.
- [131] H. Zheng, J. Yao, Y. Zhang, and I. W. Tsang. Degeneration in vae: in the light of fisher information loss, 2018.

Appendix A

Hyperparameters and Supplementary Plots

A.1 Hyperparameters (Chapter 2)

Hyperparameters for SAFESPAM, CHUNC and CHUNC2

Table A.1: Table of Hyperparameters for SAFESPAM. Note for pMSSM m_H trial 1 training was split between 20 epochs with a proportional mix of valid and invalid data and 20 epochs of only valid data.

Space	cMSSM	cMSSM	pMSSM	pMSSM	pMSSM
Constraint	m_H	Ω_{DM}	m_H (Tr. 1)	m_H (Tr. 2)	Ω_{DM}
Batch Sz.	2500	2500	2500	2500	2500
L.R.	0.0005	0.0005	0.0005	0.0005	0.0005
Opt.	ADAM	ADAM	ADAM	ADAM	ADAM
num. layers	10	10	10	10	12
Act.	Leaky RELU	Leaky RELU	Leaky RELU	Leaky RELU	Leaky RELU
num. epochs	60	6160	40*	160	160
α	5	5	5	5	5
β	1	5	1	1	1
γ	0	0.1	0.3	0.3	0.8

Table A.2: Table of Hyperparameters for CHUNC and CHUNC2.

Model	CHUNC	CHUNC2	CHUNC	CHUNC2
Space	cMSSM	cMSSM	pMSSM	pMSSM
Constraint	Triple	Triple	Triple	Triple
Batch Sz.	64	64	64	64
L.R.	0.001	0.001	0.001	0.001
Opt.	ADAM	ADAM	ADAM	ADAM
num. layers	22	22	22	22
Act.	Leaky RELU	Leaky RELU	Leaky RELU	Leaky RELU
num. epochs	100	100	100	100
α	1	1	1	1
β	1	1	1	1
γ	1	1	1	1
c	1	N/A	1	N/A
d	1	N/A	1	N/A

A.2 Hyperparameters (Chapter 3)

Table A.3: Table of Hyperparameters

Num. Transforms	4
Batch Sz.	$\begin{cases} 25000, & T_i > 25000 \\ 250 & T_i \leq 25000 \end{cases}$
Num. Epochs	600
B	3
NN Hidden Features	100

A.3 Supplementary plots (Chapter 3)

A.3.1 Valid points beyond initial boundaries

While the boundaries specified in table 3.2 define the intended scope of our search, our methodology also produced a lesser number of 7-valid points which fall outside those bounds. Because these points are generated as an artifact of our iterative search procedures, they do not convey information about the relative proportionality of the valid subspace in these regions. Nevertheless we include them here.

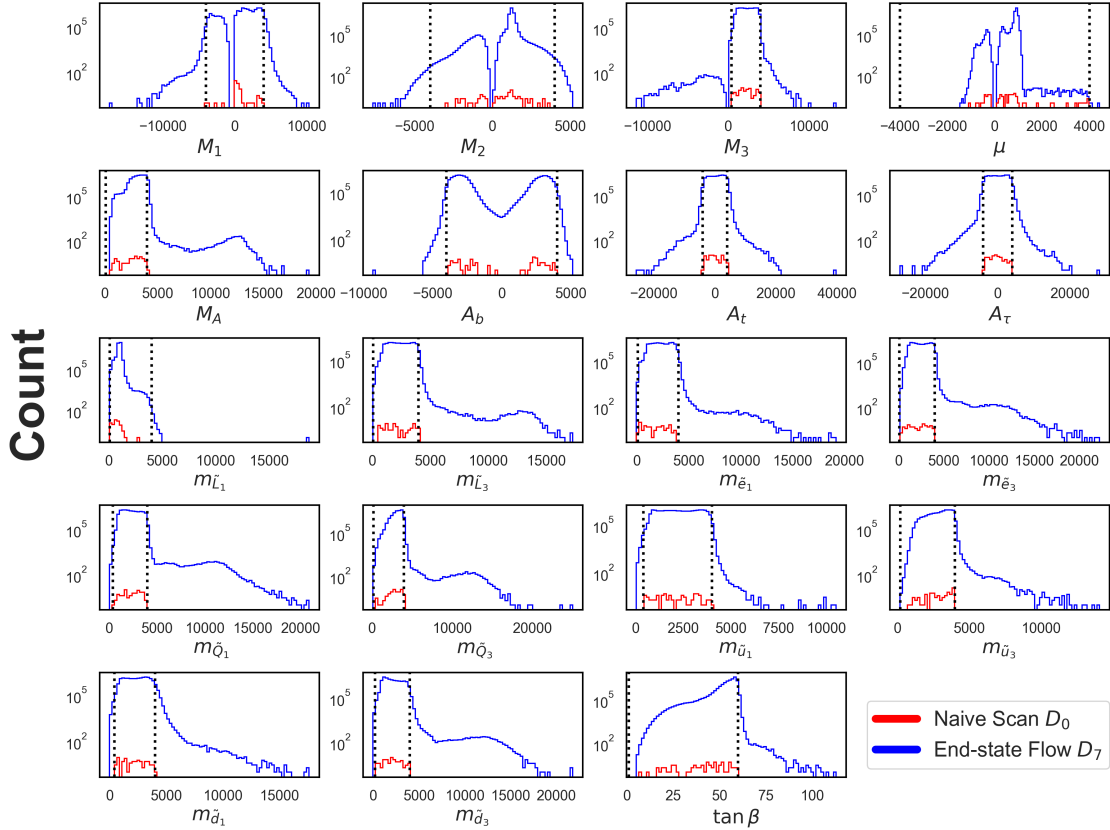


Figure A.1: The same two distributions from figure 3.5, S_0^7 from D_0 in red, and S_7^7 from D_7 in blue, are shown here, but with the axes expanded to include the 7-valid points found beyond the initial scan's boundaries, which are represented via dashed lines.

A.3.2 Additional 2D Histograms (Chapter 3)

In this work we focus primarily on displaying our results via 1D histograms. While it would be prohibitive to include all possible unique 2D Histograms of the same results, we recognize that some common choices may be useful for comparison to future studies, or show interesting proportionalities of the 7-valid manifold, and include them here.

A.4 Hyperparameters (Chapter 4)

Table A.4: Table of hyperparameters for generating correctional dataset and smart iteration dataset

Correctional Dataset	cMSSM	pMSSM
LOKV size	$\mathcal{O}(187.5k)$	$\mathcal{O}(9.84M)$
Num. KDE dummy pts	300k	300k
Num. weak targets per GUT pt	50	50
Num. total ordered pairs	15M	15M
σ_c	$\frac{1}{48}$	$\frac{1}{48}$
Smart Iteration Dataset	cMSSM	pMSSM
Num. weak targets per GUT pt	25	N/A
σ_s	$\frac{1}{48}$	N/A
Num. valid duplicates	25	N/A
Num. total ordered pairs	$\mathcal{O}(9.38M)$	N/A

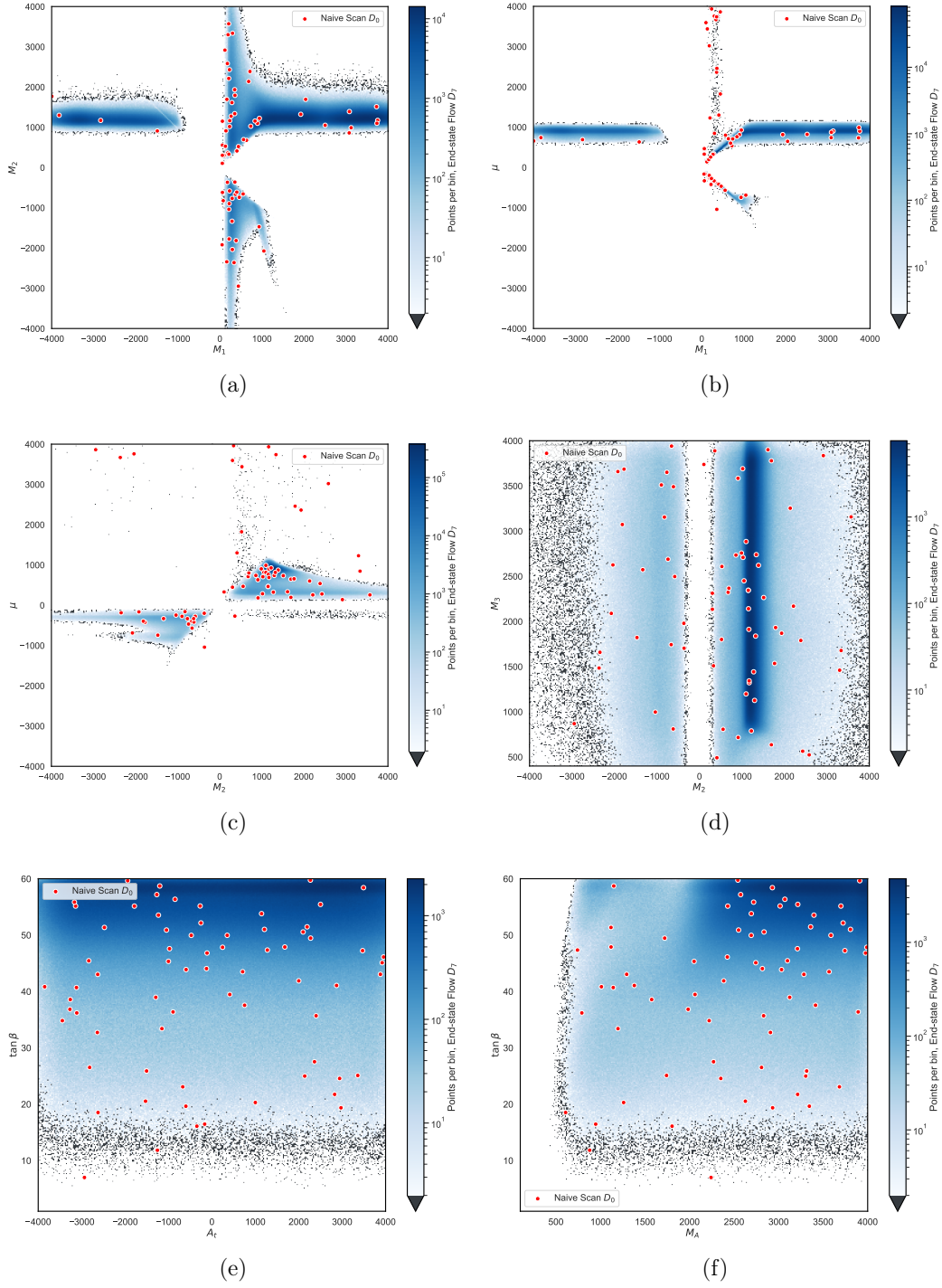


Figure A.2: Selected 2D histograms comparing the 7-valid distributions S_0^7 from D_0 in red, and S_7^7 from D_7 in blue. Bins containing only a single theory point from S_7^7 are shown in charcoal for improved visibility in sparse regions and the axes are set to exclude points outside the initial scan boundaries. The parameters plotted are as follows: a.) M_2 vs. M_1 b.) μ vs. M_1 c.) μ vs. M_2 d.) M_3 vs. M_2 e.) $\tan \beta$ vs. A_t f.) $\tan \beta$ vs. M_A

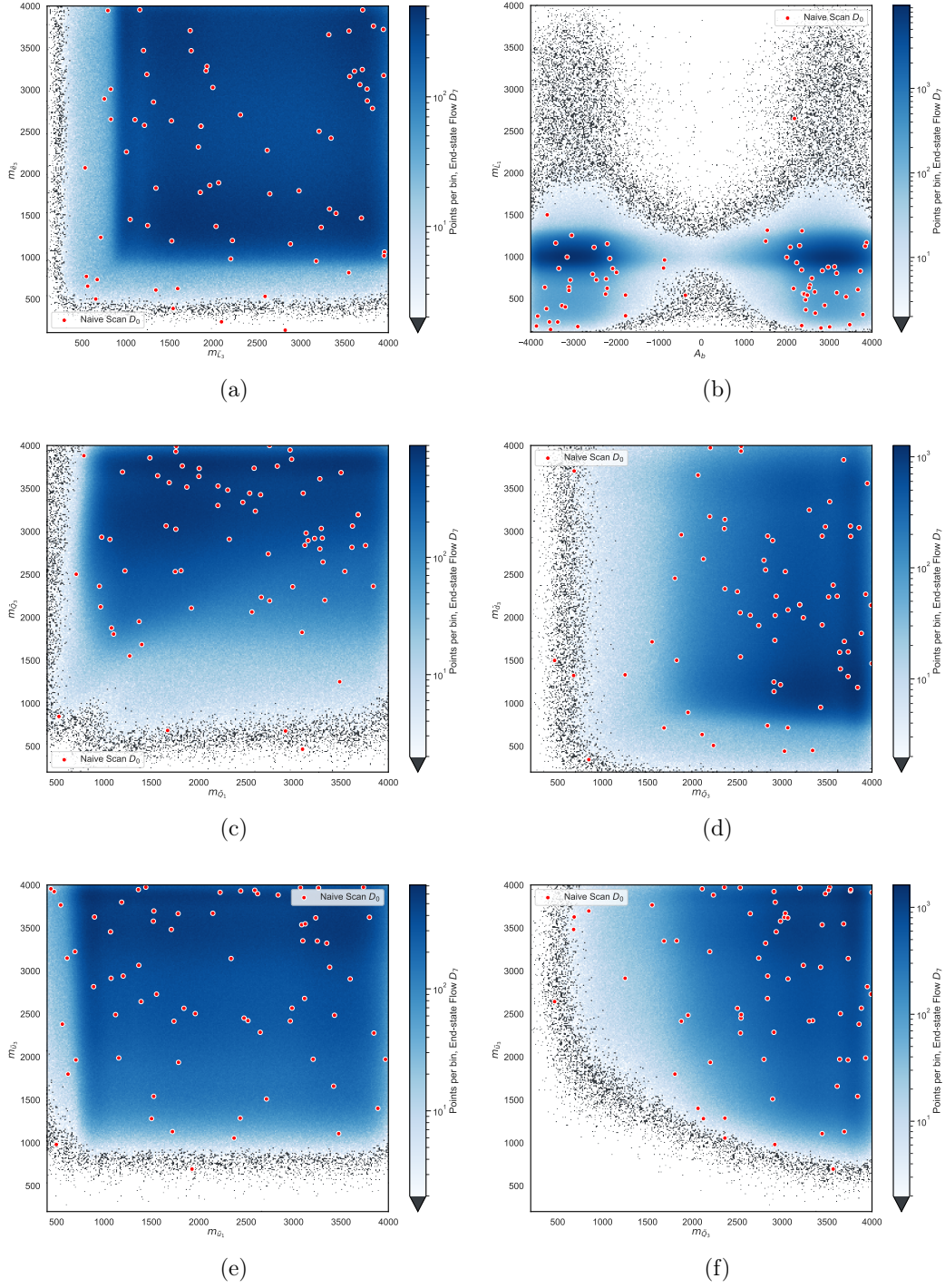


Figure A.3: 2D histograms comparing the 7-valid distributions S_0^7 from D_0 in red, and S_7^7 from D_7 in blue. Bins containing only a single theory point from S_7^7 are shown in charcoal for improved visibility in sparse regions and the axes are set to exclude points outside the initial scan boundaries. The parameters plotted are as follows: a.) $m_{\tilde{e}_3}$ vs. $m_{\tilde{L}_3}$ b.) $m_{\tilde{L}_1}$ vs. A_b c.) $m_{\tilde{Q}_3}$ vs. $m_{\tilde{Q}_1}$ d.) $m_{\tilde{d}_3}$ vs. $m_{\tilde{Q}_3}$ e.) $m_{\tilde{u}_3}$ vs. $m_{\tilde{u}_1}$ f.) $m_{\tilde{u}_3}$ vs. $m_{\tilde{Q}_3}$

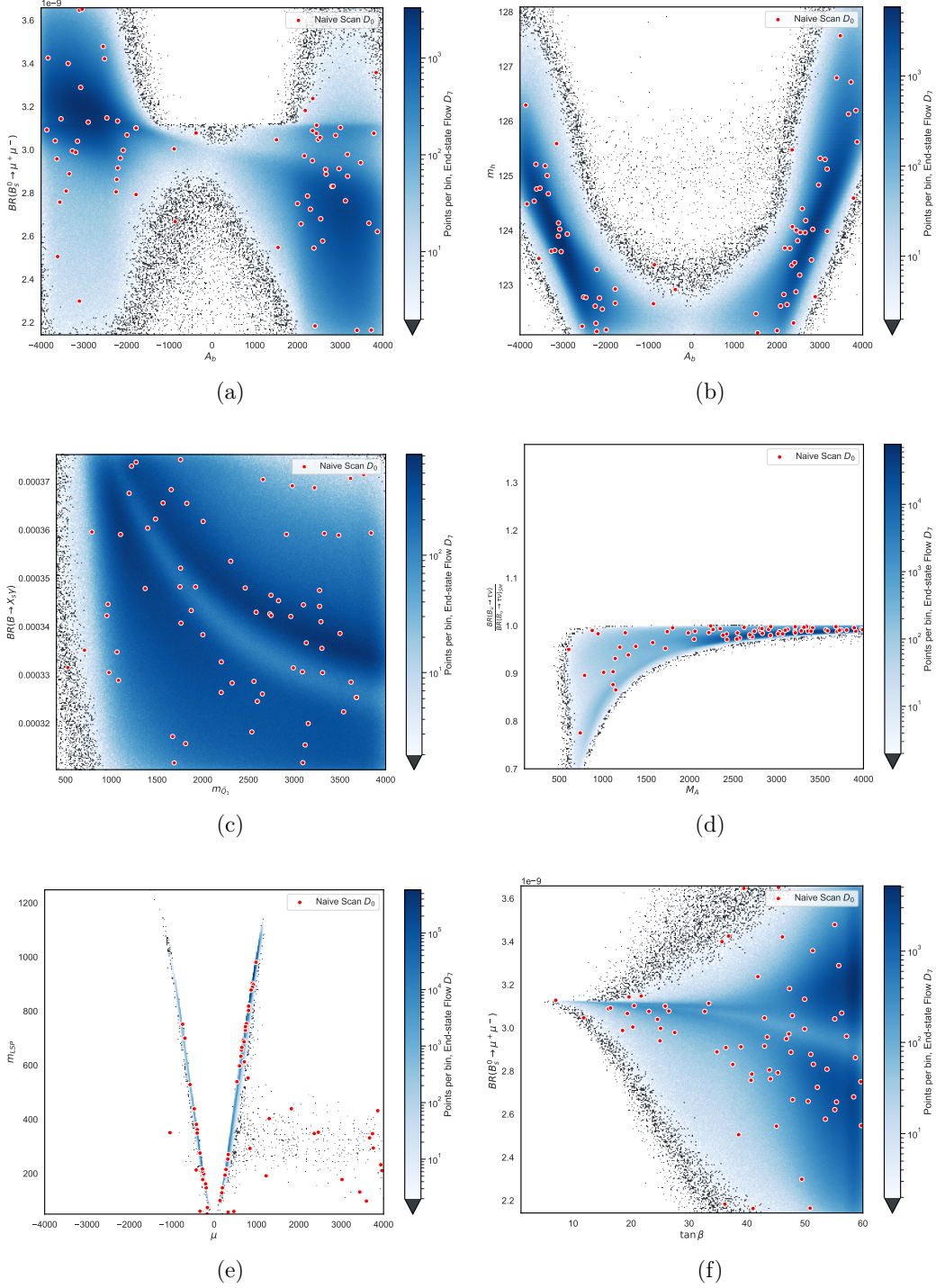


Figure A.4: Selected 2D histograms comparing the 7-valid distributions S_0^7 from D_0 in red, and S_7^7 from D_7 in blue. Bins containing only a single theory point from S_7^7 are shown in charcoal for improved visibility in sparse regions and the axes for GUT-scale parameters are set to exclude points outside the initial scan boundaries. The parameters plotted are as follows: a.) $BR(B_s^0 \rightarrow \mu^+ \mu^-)$ vs. A_b b.) m_h vs. A_b c.) $BR(B \rightarrow X_s \gamma)$ vs. $m_{\tilde{Q}_1}$ d.) $\frac{BR(B_u \rightarrow \tau \nu)}{BR(B_u \rightarrow \tau \nu)_{SM}}$ vs. M_A e.) m_{LSP} vs. μ f.) $BR(B_s^0 \rightarrow \mu^+ \mu^-)$ vs. $\tan \beta$

Table A.5: Table of hyperparameters for training correctional and smart iteration networks. The cMSSM correctional network was trained switching optimizers partway through: the SGD [116] portion of the training uses a learning rate 10x the one used for ADAM. The smart iteration network uses a Leaky ReLU as activation function, but the final output is passed through a hard sigmoid. Check-pointing was used in some instances to select models with the best test loss performance during training.

Correctional Network	cMSSM	pMSSM
Batch Sz.	5000	150
L.R.	0.0005 (ADAM) 0.005 (SGD)	0.0008
Opt.	ADAM (120 epochs) SGD (80 epochs)	ADAM
num. layers	8	8
Act.	Leaky ReLU	Leaky ReLU
num. epoch	200 total	35
Smart Iteration Network	cMSSM	pMSSM
Batch Sz.	10000	N/A
L.R.	0.00005	N/A
Opt.	ADAM	N/A
num. layers	8	N/A
Act.	Leaky ReLU*	N/A
num. epoch	1472	N/A