Symmetry discovery with deep learning

- Krish Desai ,Benjamin Nachman and Jesse Thaler

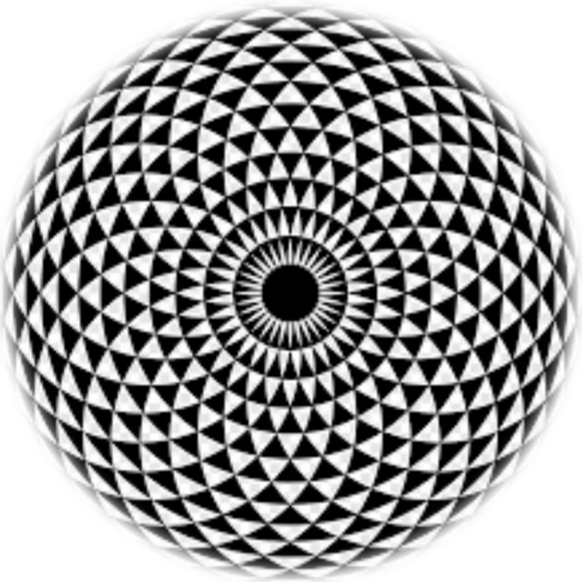
A review by Rajavivegan [NA22B085]

Overview of Symmetry

Symmetries are embedded in every aspect of physics and data in general. Symmetries refers to invariance under certain transformation. Symmetries have been the fundamental reason we have various conserved quantities and invariants which helped us to analyse physics of various systems with relative ease. Symmetries reduces the complexity of analysing systems.

There are a variety of definitions for symmetry ranging from intuitive ones to rigoruous ones.

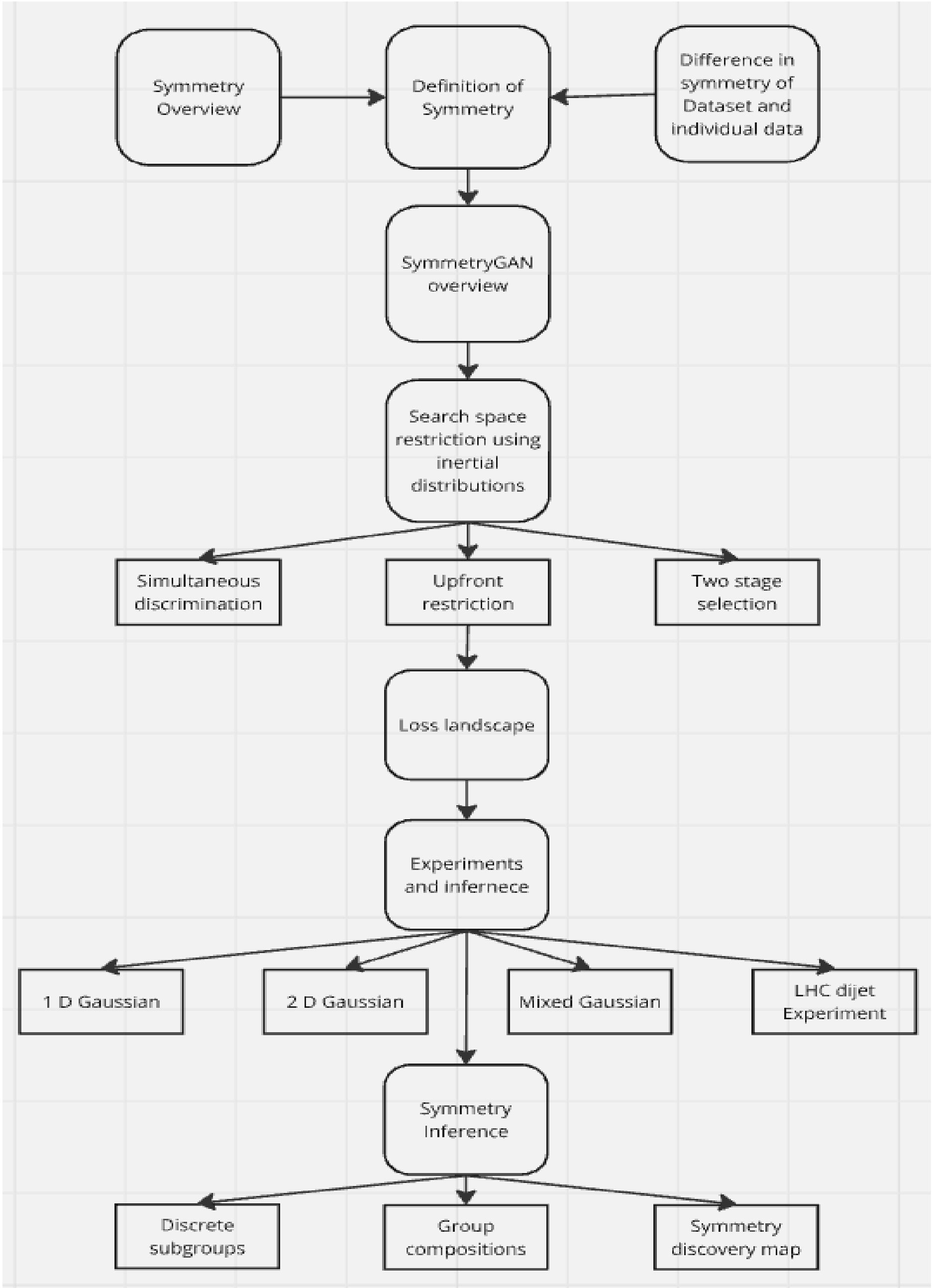
We try to present the various definitions and views before diving into deep learning implementations



Problem Addressed

Neural networks are proven to be great in capturing patterns from higher dimensional data and hence this research paper presents a way to find symmetries in datasets using their novel deep learning architecture 'SymmetryGAN' which is a modified Generative Adversarial Neural network (GAN).

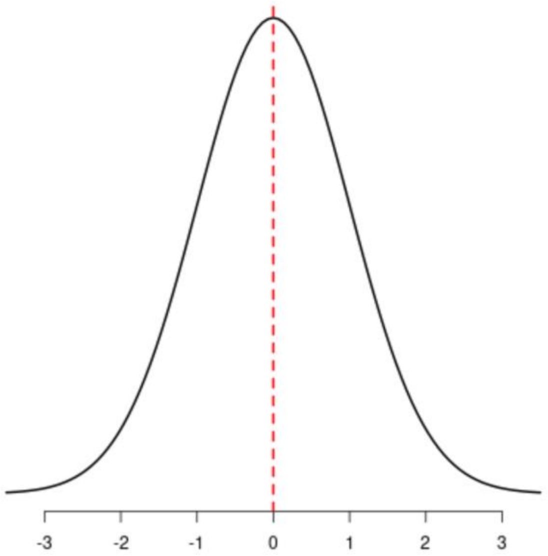
The paper clearly states that it attempts to find symmetries in datasets as a whole and in individual data. This kind of symmetries help to create synthetic datasets, revolutionizing simulations and addressing data scarcity.



A flow map of the research paper

Symmetry of datasets and individual data

- Symmetry of individual data is different from symmetry of a dataset
- Let X be a random variable and x be its instantiation. Then a transformation *h(x)* is said to be symmetry of a individual data if *h(x)=x*
- It could also be such that the transformation leaves preserves certain properties or functions unchanged.
- We don’t deal with this symmetry.



- Symmetry of the dataset refers to a transformation that leaves the statistical properties of the dataset unchanged.
- This means that an instant drawn from the original and transformed dataset should be statistically identical.
- Such symmetries give us a way to perform data augmentation efficiently.

Such transformations are the symmetries we intend to find. We parameterize those transformations using insights from group theory and learn those parameters using the architecture proposed

Symmetry – A statistical definition

- Dataset is treated as a statistical distribution.
- A symmetry is defined as a map g such that it preserves the pdf of the distribution. Mathematically,

$$p(X = x) = p(X = g(x))\,|g'(x)|$$

$|g'(x)|$ is the Jacobian of the transformation.

- Jacobian quantifies how much the transformation compresses or stretches the space at each point.
- Jacobian in the equation makes sure that the transformation preserves the property that integration of pdf over entire space is 1.

- Consider a uniform distribution in R $F(X) \sim \mathcal{U}[0, 1]$, . A transformation given by $\tilde{g}(x) = 1 - x.$ is also a uniform distribution in [0,1]. The pdf of the original variable and the transformed variable are same
- Hence it is said to be a symmetry of the dataset.

- For any distribution f(X), the cumulative distribution function F(X) is also a random variable given by \ $F(X) \sim \mathcal{U}[0, 1]$, is a uniform distribution
- This property is used in systems to simulate random variable of from a given distribution by first simulating a uniform random variable and applying inverse which gives out a random variable following the distribution.
- From this we can say that every distribution has a pdf preserving map $g = F^{-1} \circ \tilde{g} \circ F$ where ° is the composition.
- For multi dimensional random variables, this transformation has to be applied to each varible seperately

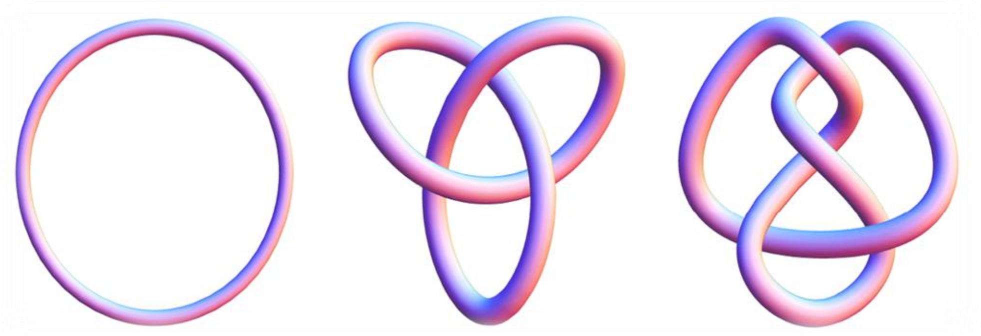
Using the above, synthetic symmetric data in 1 dimension can be generated which is technically a symmetry discovery but it is a quantile map, meaning that it is different for each variable of a multidimensional system and has to be applied seperately to each variable.
In physics,such quantile maps are not considered symmetries.

In physics we need a map that can be more generalized and robust such that it can be extended for similar systems.

Hence there is a need for redefining the definition of symmetry

Symmetry – A rigorous definition

- Symmetry is mathematically studied under group theory.
- A group is a collection of objects in a space with a group operation that satisfies 4 properties – closure, inverse, identity and associativity.
- A symmetry can be imagined as a set of transformations on an object satisfying the 4 properties. It could be rotation, translation or some higher dimensional complex transformations.
- Groups are of various kinds based on the space/object the symmetry operates on and a few of them are of our interest here.
- Groups can be discrete or continuous.



- A knowledge in group theory is essential to come up with parametric equations for symmetry transformations which can be learned using neural networks.
- A **symmetry group** is a group whose elements are the symmetry transformations of a particular object or system, and the group operation is the composition of these transformations.

General Linear groups and its subgroups

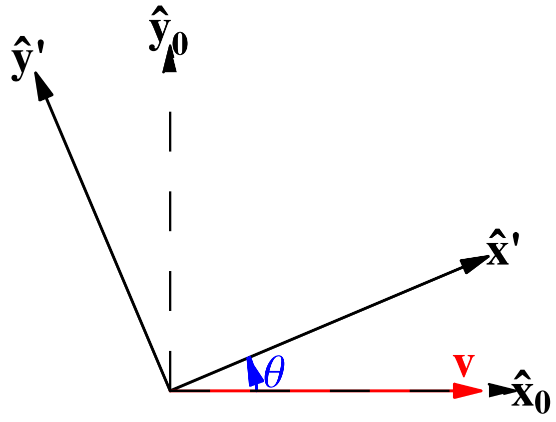
The general linear group: $GL_n(R)$ consists of all invertible nxn matrices over the real numbers . It includes all possible linear transformations in a N-dimensional space, such as rotations, reflections, scalings, and shears.
•This is a vast group and has well defined subgroups which are generally of our interest when it comes to symmetry

The orthogonal group : O(n) consist of all the nxn matrices M such that its inverse is its transpose. The matrices O(n) preserve the length of vectors and are thus related to distance-preserving transformations like rotations and reflections.

$$M^T M = I$$

- **The special orthogonal group:** denoted SO(n), is a subgroup of the orthogonal group O(n) and consists of all orthogonal matrices with determinant +1. Matrices in SO(n) represent rotations but do not inlcude a reflection (reflection requires the determinant to be -1)

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$ Is the SO(2) group which is the rotation matrix in 2 dimension



The affine group: $Aff_n(R)$ consists of all transformations that can be written as a combination of a linear transformation (rotation, reflection, scalings and shear) and a translation. An affine transformation has the form

$X'=AX+B$ .where X' is the transformed space
A is a invertible matrix from $GL_n(R)$ group
B is the translation vector

- Another group of interest is the $ASL_n^{\pm}(R)$ which is a subgroup of the affine group where the determinant of A is restricted to be ±1. It has n(n+1) parameters
- Determinant of the transformation matrix being ±1 means that the transformation is volume preserving.

Abelian group: a group where the group operation is commutative, meaning for any two elements a and b in the group, a·b=b·a

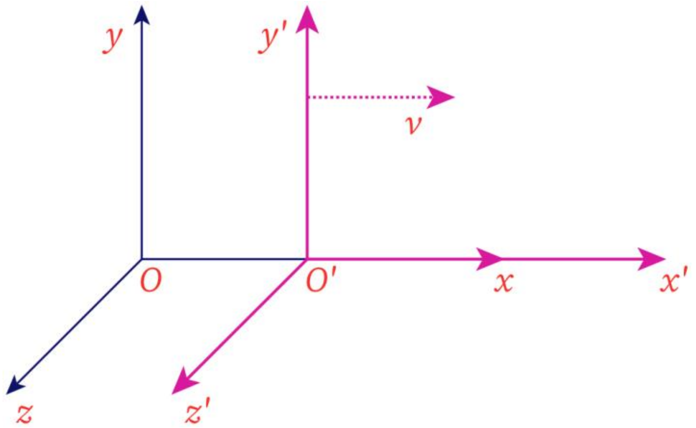
•The important Abelian group for our study is the cyclic group denoted by where application of transformation n times leaves the system unchanged Z_n
•The rotation of a 2D space by π two times leaves the space unchanged. Hence it's a cyclic group Z_2

Innertial distribution

- In physics, every measurement is relative and so is symmetries. Hence there is a need to embed relativeness in the definition of symmetry .
- Consider two spaces $\mathbb{R}^n$ with distributions given by $p: \mathbb{R}^n \to \mathbb{R}_{\geq 0}$, $p_I: \mathbb{R}^n \to \mathbb{R}_{\geq 0}$,

  A map $g: \mathbb{R}^n \to \mathbb{R}^n$ is defined to be a symmetry of p relative to p_I if it is PDF-preserving for both p and p_I . That is,

$$p(x) = p(g(x))|g'(x)|, \qquad p_I(x) = p_I(g(x))|g'(x)|$$



- The reference density is called the inertial density and it's a analogue to inertial frame of reference.
- Since in physics, most of the problems are set in Eucledian space , a uniform distribution in $\mathbb{R}^n$
  is a canonical choice.
- This also narrows down our search space as we will see in the upcomming slides.

Inertial restriction and implementation

- The condition also requires that the likelihood ratio which is the ratio of pdf of the two distributions before and after transformation remains the same

$$\ell(g(x)) = \frac{p(g(x))}{p_I(g(x))} = \frac{p(x)}{p_I(x)} = \ell(x).$$

- The choice of uniform distribution in $\mathbb{R}^n$ implies that the symmetry map g should be an Equiareal map.
- An Equiareal map is the one that preserves area of curves.
- To see why this is true we have to realize that for a uniform distribution defined in a open set $\mathcal{O} \subseteq \mathbb{R}^n$, we have

$$p(x) = \frac{1}{\mathrm{Vol}(O)} \quad \text{for all } x \in O,$$

Where Vol(O) is the volume under the distribution which is same for both p(x) and p(g(x)).

- This naturally implies that the jacobian should be 1 : The distances in space remains the same and so is the area of surfaces

Restricting the search space using Equiareal maps

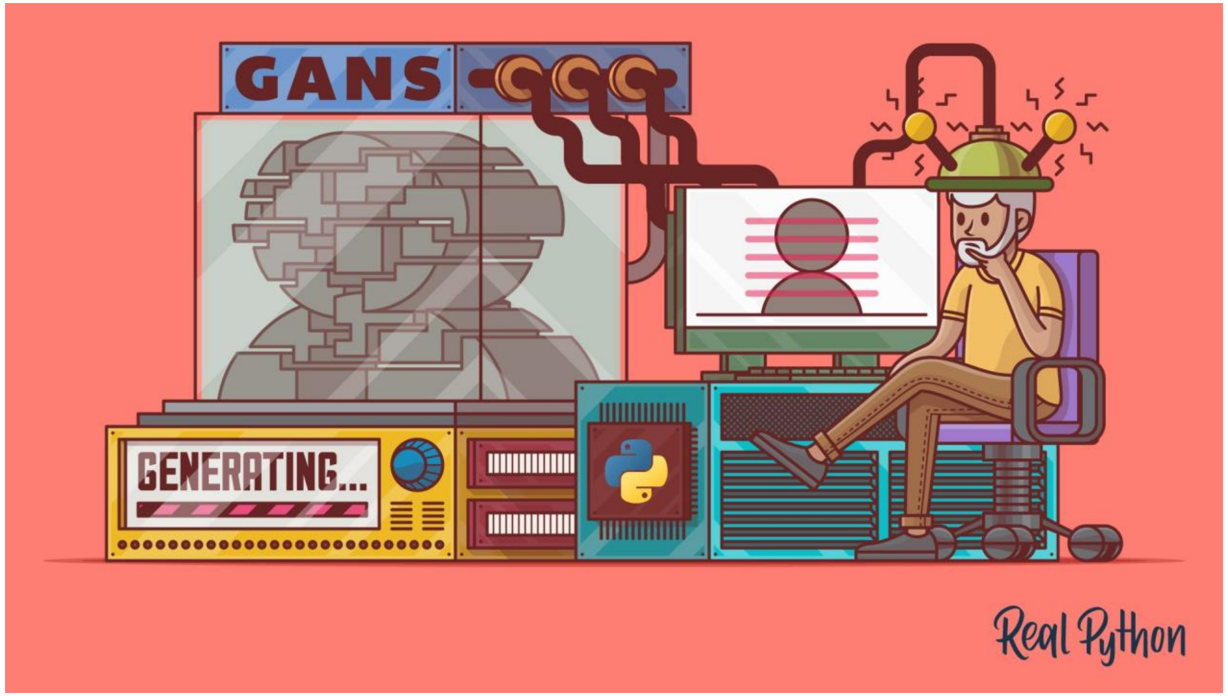
- Equiareal maps could be linear or non linear for every space of dimension n>1.
- One such example of non linear map in n=2 is the Henon map given by g(x,y)=(x,y-x^2). We see that its Jacobian determinant is 1 everywhere.
- Since non linear maps are very hard to parametrize, the paper restricts studies to only linear maps.
- More specifically the $ASL_n^{\pm}(R)$ which is defined previously is a group of linear transformations and thus has linear symmetries.
- This makes parametrization of the problem easy as this has complete parametrization, one of which is the Iwasawa decomposition.

$$\mathbb{A}SL_n^{\pm}(\mathbb{R}) = \{g\colon \mathbb{R}^n \to \mathbb{R}^n | g(x) = Mx + V,$$
$$M \in \mathbb{R}^{n\times n}, \det M = \pm 1, V \in \mathbb{R}^n\}.$$

Iwasawa Decomposition: This is a technique from the theory of Lie groups that allows one to decompose a matrix into a product of three specific matrices—one that is lower triangular, one that is orthogonal, and one that is upper triangular with positive diagonal entries.

The Adversarial model

- Now that we have seen the mathematical requisites and definitions needed to parametrize the problem, lets now focus on the implementation. The core innovation of the paper is the introduction of *SymmetryGAN*, a generative adversarial network (GAN)-based framework designed to discover symmetries in datasets



- The Generative Adversarial Neural network (GAN) is a novel unsupervised learning approach for generating new data They find their application in various image generation software.
- They are great in capturing the latent space and the learning mimics natural evolution of two species competing for resources and arriving in a stable equilibrium.
- A standard GAN consists of two networks:
  - **Generator** : Tries to generate data that resembles the true data distribution. Random noise data is fed into generator and it captures the latent space in the noise while trying to fool the discriminator.
  - **Discriminator** :It is a regular classifier which attempts to distinguish between real data and data generated by the generator.
- The two networks are trained in a minimax game where the generator tries to fool the discriminator by generating realistic data, and the discriminator tries to correctly identify the generated data as fake.The training ends when the discriminator can't distinguish between the original data and fake data giving equal probability i.e 0.5 to every data it encounters.

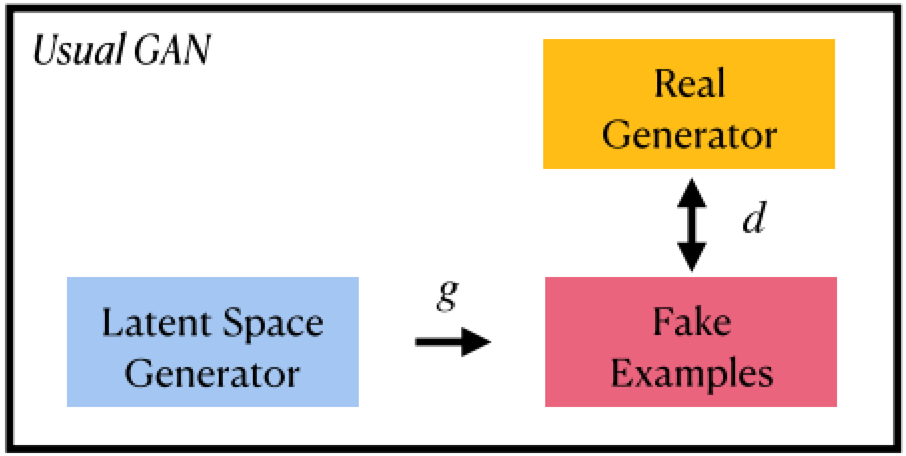
The SymmetryGAN

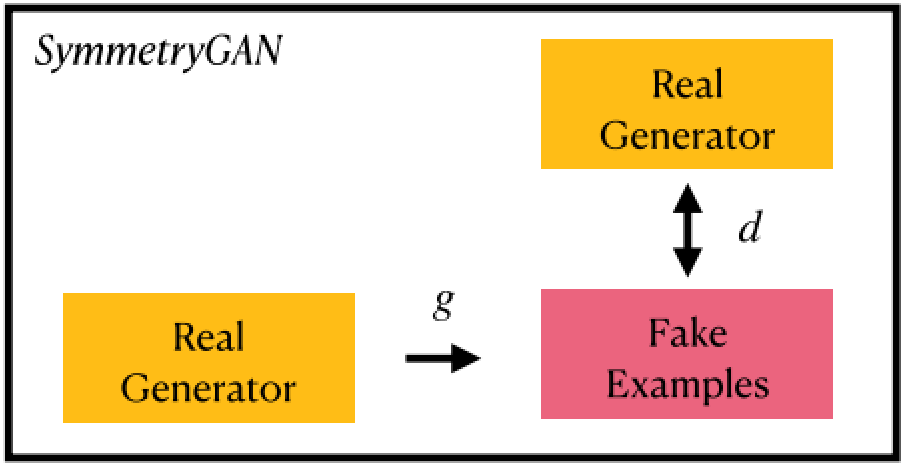
- The symmetryGAN proposed in this paper resembeles standard GAN in every aspect except that the original data is fed into the generator and it applies transformation to it.
- If the discriminator is unable to distinguished the original and trasformed data, then it is inferred to be a Symmetry of the system. Thus there is simultaneous learning of the two functions g and d

$$g: \mathbb{R}^n \to \mathbb{R}^n,$$

$$d: \mathbb{R}^n \to [0, 1].$$

- The generator is randomly initialized on the search manifold, and through gradient descent, converge to the nearest symmetry
- The dissadvantage is mode collapse which is when the generator converges to a identity map





Implementing Inertial restriction

- Neural Network architectures can be designed such that they respect certain symmetries. The translational symmetry embedded in the structure of CNN is one such example
- To implement the inertial restriction, the authors propose 3 solutions

	Simultaneous discrimination	Two stage selection	Upfront restriction
Description	• In this method, the discriminator is tasked with distinguishing both the input dataset and the inertial distribution simultaneously. • The **discriminator** is applied to two types of data during training:	• The generator first identifies PDF-preserving maps for the dataset, and then a post hoc selection is made to filter out the maps that also preserve the inertial distribution. • Same as simultaneous discrimintion but done in two steps	• The search space for the generator is **restricted upfront** to include only transformations that are known to preserve the inertial distribution. For example, if the inertial distribution is uniform, you could restrict the generator to equiareal maps.
Advantages	This approach doesn't require knowledge of the symmetries of the inertial distribution upfront.	Computationally straightforward and has same advantages as that of simultaneous discrimination	Computationally efficient as it narrows down the search space
Disadvantages	Requires sampling from both the real dataset and the inertial distribution along with discriminator handling 2 datasets which increases computational complexity.	Highly inefficient and **computationally wasteful** because the space of PDF-preserving maps for the dataset is generally much larger than the space of symmetry maps that also preserve the inertial distribution.	Semi automated and requires us to have a prior knowledge over the pdf preserving maps of inertial distribution.

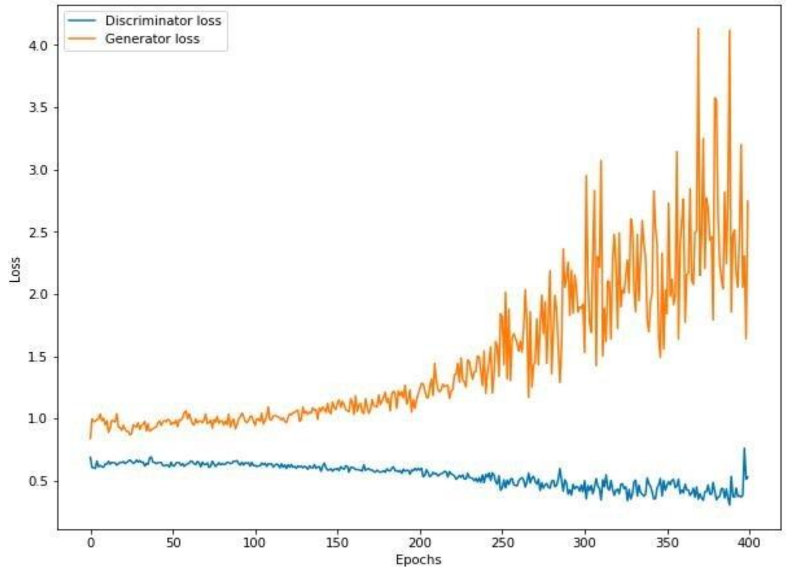
The cross entropy loss function

- The loss function of the SymmetryGAN is similar to the binary cross entropy loss function of a regular GAN. The discriminator and the generator are parametrized as neural networks. It is given by

$$L[g,d] = -\frac{1}{N}\sum_{x\in\{x_i\}_{i=1}^N}[\log(d(x)) + \log(1-d(g(x)))].$$

.d(x) is the probability (output) of the discriminator

- The generator tries to maximize this loss by minimizing the second term while the discriminator tries to minize it resulting in a Min-Max game.
- The optimal d for a fixed g is given by $d_* = \dfrac{p(x)}{p(x) + p(g(x))|g'(x)|}$



- When the system learns symmetry i.e $p(x) = p(g_*(x))|g_*'(x)|$, d* becomes 0.5 and the loss function becomes 2log2. This condition is very helpful in symmetry inference which is discussed in later slides.

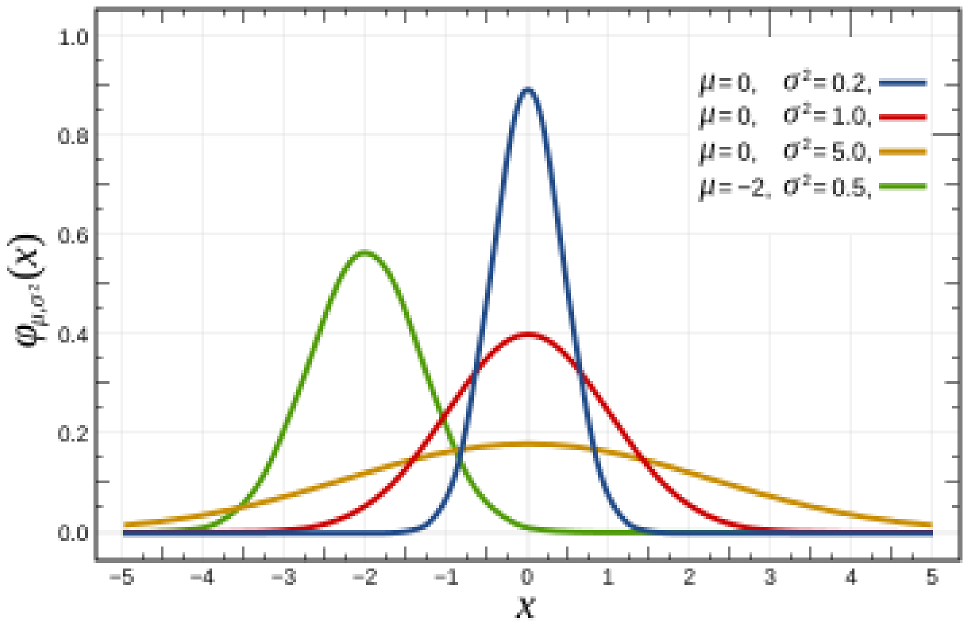
- To target a particular symmetry from a subgroup of general linear group , we can add additional loss terms.
- cyclic symmetry group, we can incorporate the additional term

$$-\frac{\alpha}{N}\sum_{x\in\{x_i\}_{i=1}^N}(g^q(x)-x)^2,$$

Which is a mean squared error term and α factor that quantifies its significance.

1D Gaussian Experiment

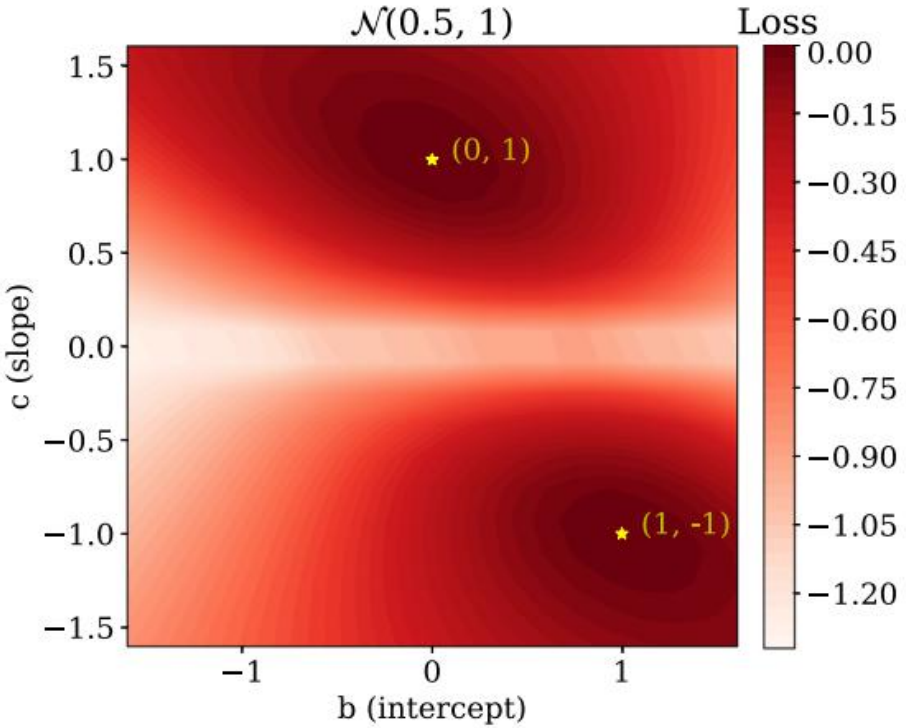
- 1 D gaussian distribution or normal distribution denoted by N(μ,Ω) where μ is the mean of the distribution and Ω is the variance
- It is a Bell shaped curve.
- The distribution considered here is N(0.5,1). The curve is centred around 0.5.



- In 1 dimension, the equiareal maps are parametrized as $g(x) = b + cx.$

- By visual inspection, we see that it has only two symmetries, one being identity and another given as 1-x

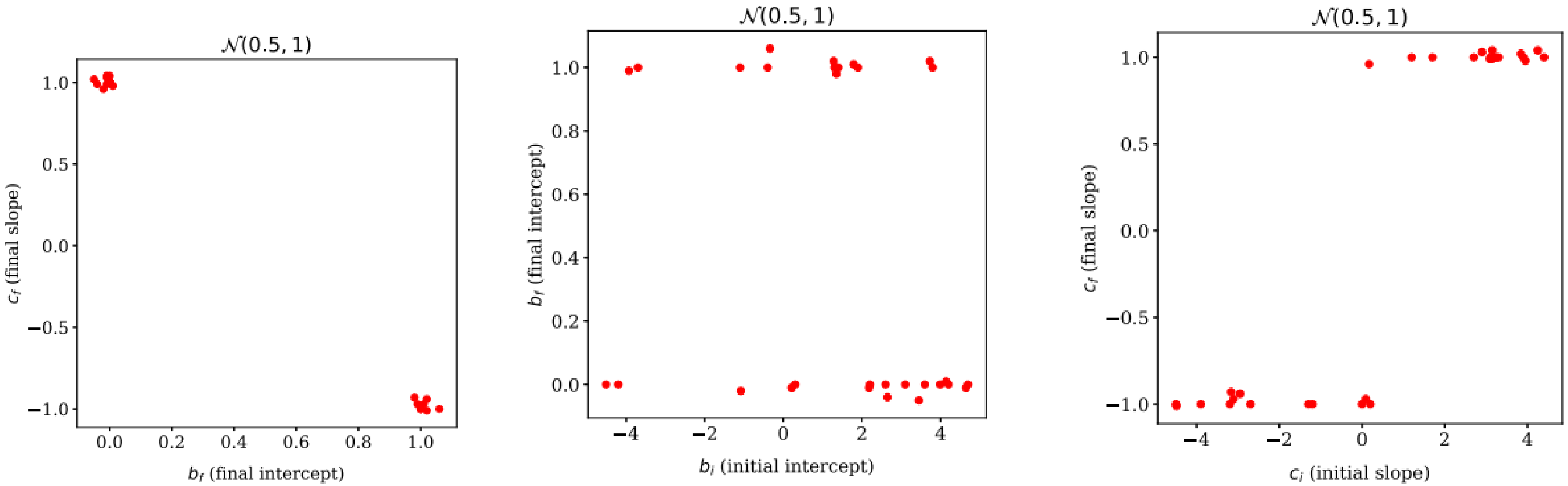
$$g(x) = x, \qquad g(x) = 1 - x.$$



- We can see this analytically with a heat map of variation of loss function with the parameters c and b.
- We inspect that the loss is maximized precisely at the two symmetries

- Since gradient descent converges to local maxima/minima, the convergence to one of the symmetries depends on the initialization of b and c .
- It is inspected from the loss landscape itself that the learned values depends only on c – the slope and not the intercept b.
- The above claim is supported by the emperical study where the initial values and the final learned values of parameters are plotted. The initial values are sampled from U2[-5,5] which is a two dimensional uniform distribution.
- In the 3rd graph we see that when c is initialized as 0, the learned b value is indeterminate which is a demonstration of loss barrier or mode collapse

We see that Symmetry GAN was able to arrive at the symmetries all the time



2 D Gaussian Experiment

- 2 D Gaussian distribution has a bell shaped 3D curve that , from visual inspection suggests that a rotational symmetry is inherent.
- We, consider two variants. First a, distrubution with mean 0 and variance 1 given by

$$N_{1,1} \equiv \mathcal{N}(\vec{0}, \mathbb{1}_2),$$

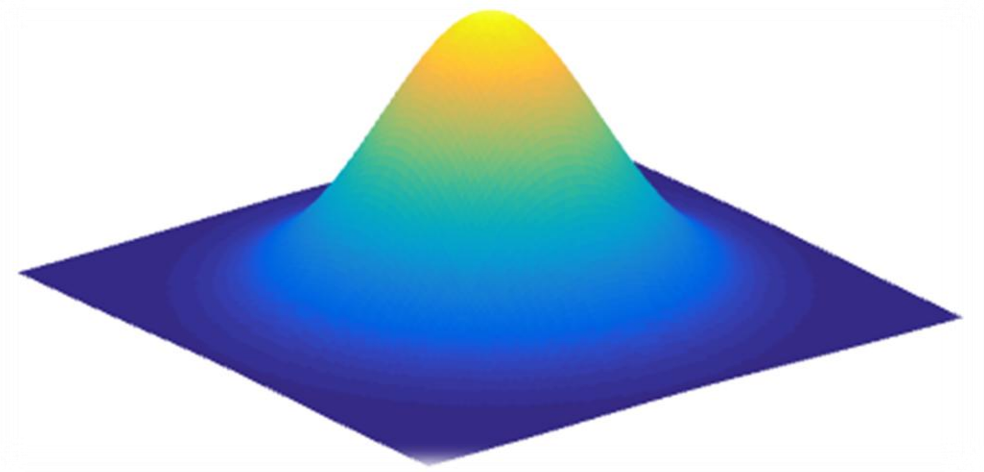
$\mathbb{1}_n$ .is a nxn Identity matrix. The non diagonals (covariance) being zero indicates that the the random variables are independent.

- It is obvious that every rotation and reflection about origin is the symmetry of this distribution. To make things a little complex we further consider another Gaussian

$$N_{1,2} \equiv \mathcal{N}\left(\vec{0}, \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}\right).$$

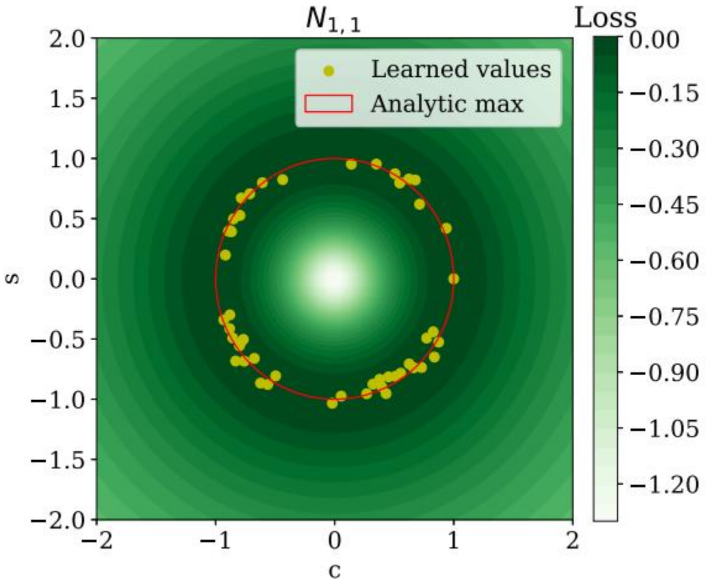
- For $N_{1,1}$: the general affine group subspace for 2D has 6 parameters. But lets consider a subgroup of symmetries – the rotations given by

$$g(X) = \begin{bmatrix} c & s \\ -s & c \end{bmatrix} X,$$

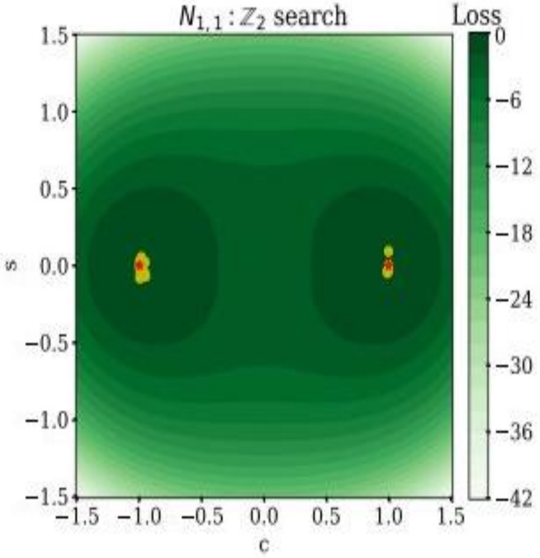


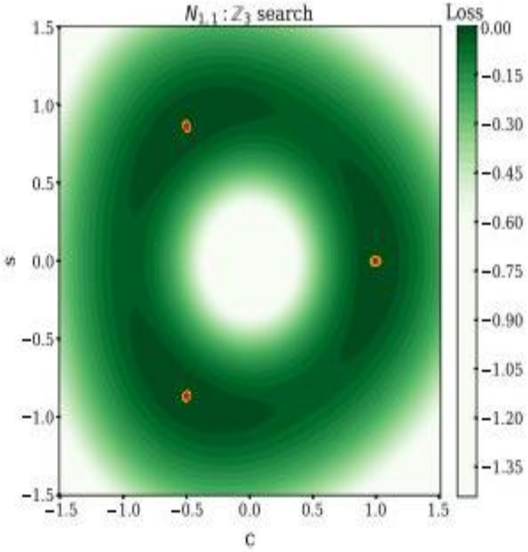
The symmetry group is given by $SO(2) \times \mathbb{R}^+ = \langle \theta, r | \theta \in [0, 2\pi), r \in \mathbb{R}^+ \rangle$.

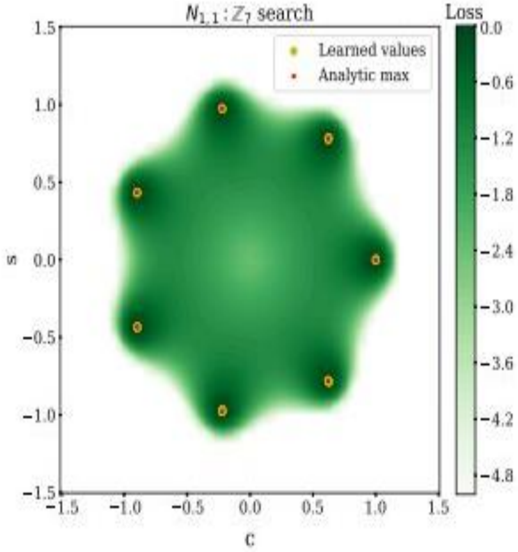
- The parameter r is considered 1 i,e the initial parameters are sampled from U[-1,1] to simplify analysis and hence the symmetry thus is the SO(2) group with a parameter theta.
- This claim is supported similarly by analytical loss landscape plot.
- For the other case, similar analysis is done and we see that the values converged to 0 and π.

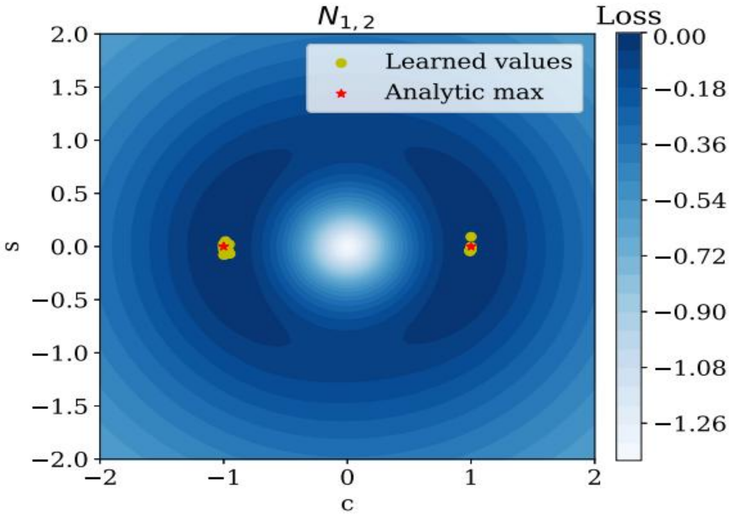


- For the $N_{1,1}$ case, since the symmerty group is continuous, it has discrete cyclic symmetric subgroups. Including the mean squared loss term to the cross entropy loss enables us to find cyclic symmetries.
- The symmetry GAN finds it accurately as it is inferred from the loss landscape for $Z_2, Z_3,$ *and* Z_7









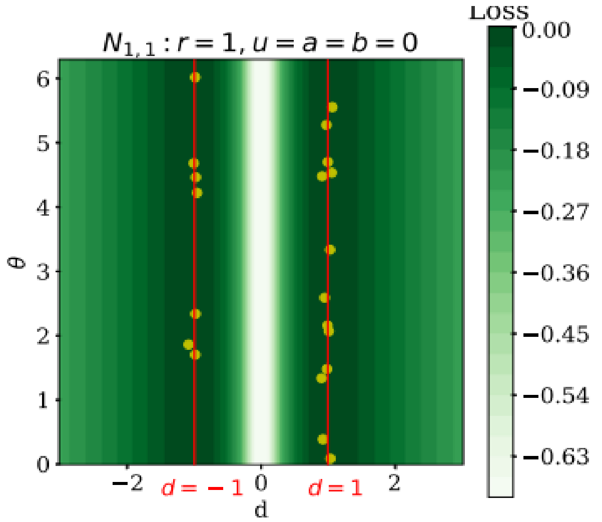
The general symmetry group of the $N_{1,1}$ is composed of 6 parameters of the $ASL_n^{\pm}(R)$ group. It is simplified by Iwasawa decomposition and for 2D case , the transformation is given by

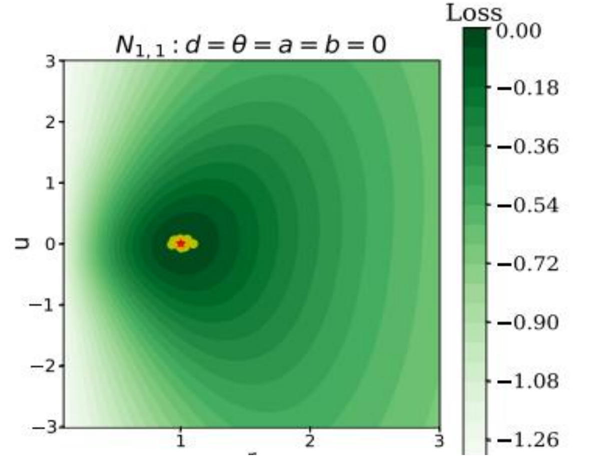
$$g(X) = \sqrt{|d|}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}^{\delta}\begin{bmatrix} c_\theta & s_\theta \\ -s_\theta & c_\theta \end{bmatrix}\begin{bmatrix} r & 0 \\ 0 & \frac{1}{r} \end{bmatrix}\begin{bmatrix} 1 & u \\ 0 & 1 \end{bmatrix}X + \begin{bmatrix} a \\ b \end{bmatrix}.$$

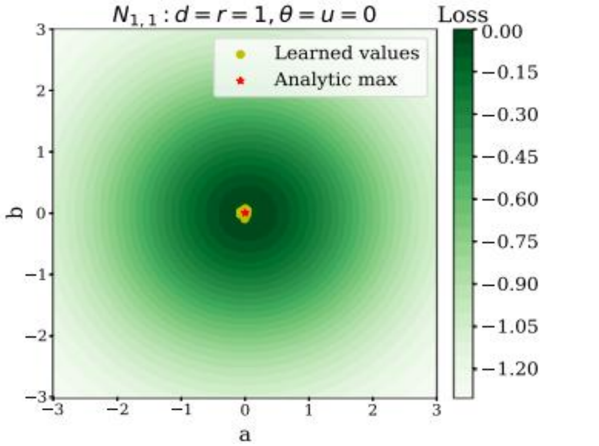
- Since it is not possible to visualize a higher dimensional space of loss landscape, slices of the plot at specific parameter values. We see that at all times SymmetryGAN converged to the actual symmetries found analytically

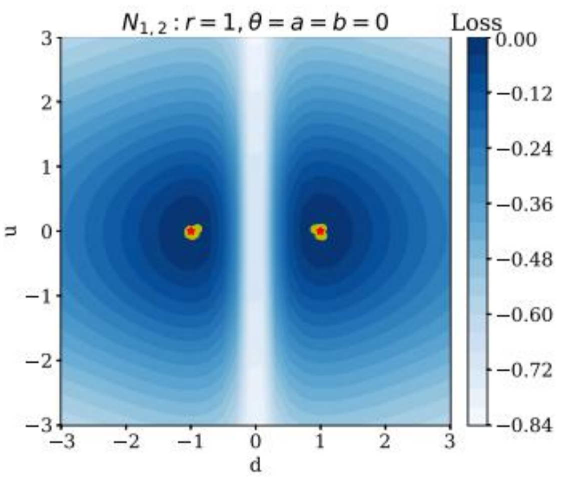
Where,

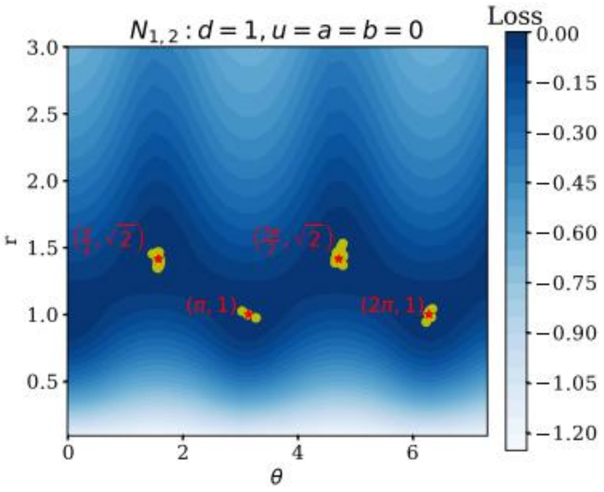
(i) $d \in \mathbb{R}^{\times}$, the determinant;
(ii) $\theta \in [0, 2\pi)$, the angle of rotation
(iii) $r \in \mathbb{R}^{+}$, the dilatation;
(iv) $u \in \mathbb{R}$, the shear in the x direction; and
(v) $(a, b) \in \mathbb{R}^2$ the overall affine shift.

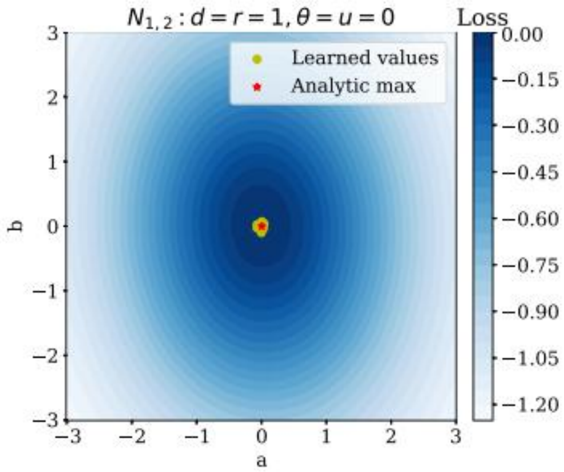












An Application in Particle physics

- The **Large Hadron Collider (LHC)** is a massive particle accelerator located at the CERN research facility near Geneva, Switzerland.
- Proton collision results in formation of intermediate particles ranging from hadrons (quarks) to hypothetical dark matter particles.
- The enormous data from LHC need to be analysed to infer insights which lead to scientific breakthroughs like the discovery of Higgs Bosons.

- Symmetries are embedded in Physics. By Noethers theorem, symmetries lead to conservation laws and the belief that these laws hold true paved way for discoveries from Einsteins mass energy equivalence to discovery of antiparticles.
- Hence there is a pressing need to infer symmetries from data.

Symmetry groups in particle physics

Coleman-Mandula Theorem:

- The theorem implies that symmetries that mix spacetime symmetries (like those in the Poincaré group) with internal symmetries (like isospin or color charge) are severely restricted. This has important implications for the structure of physical theories, particularly in particle physics, where it suggests that the only possible symmetries are those that separately affect spacetime and internal degrees of freedom.
- This states that the symmetries are subgroups of the Poincaré group.
- The Poincaré group is the semidirect product of the Lorentz group (which includes rotations and boosts) and the translation group (which includes translations in space and time). Mathematically, if x is a point in spacetime, a transformation g in the Poincaré group can be written as:

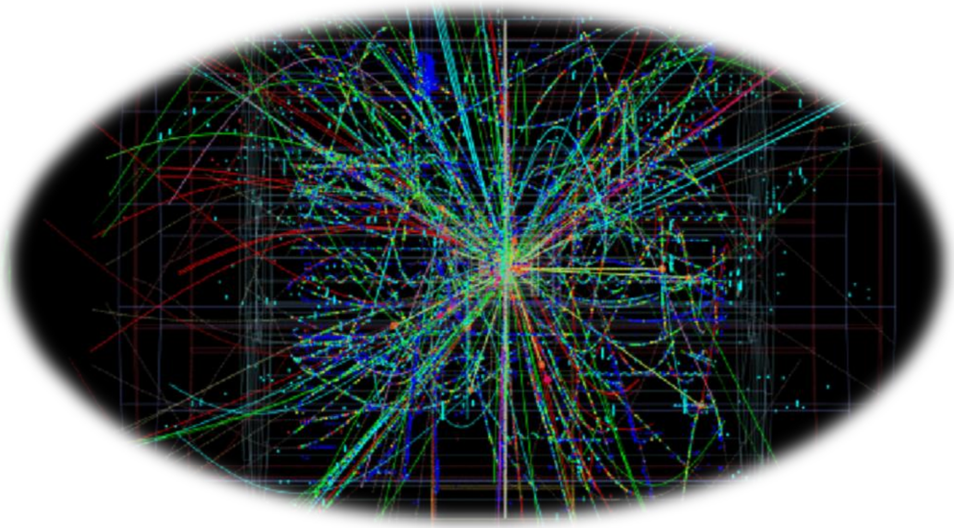
$$X' = \Delta X + a$$

Δ is lorentz transformation (rotation or a boost)
a is translation

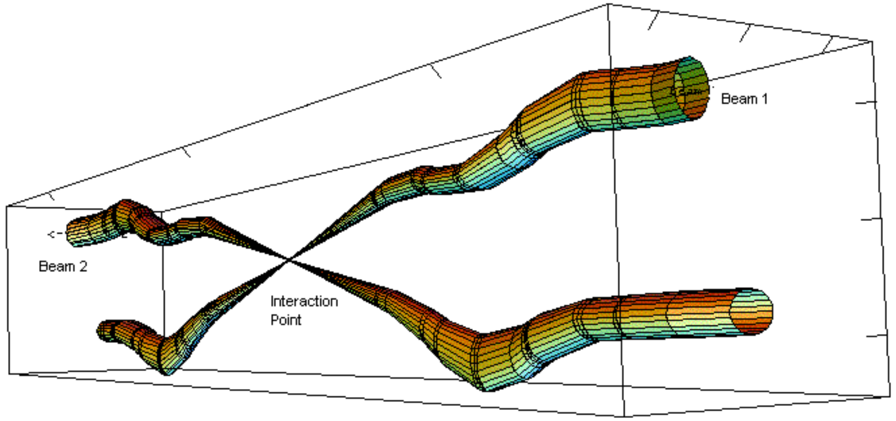
- The Poincaré group is vast and we don’t even have a representation for its classification. It has rich and complex subgroup.
- In this paper, the SymmetryGAN attempts to find symmetries within certain subgroups of Poincaré group
- The success of SymmetryGAN in finding symmetries in these subspace is a good indicator that it can find symmetries in other comples subspaces too

Dataset

- Quarks and gluons are produced in the initial collision but are not observed directly because they are confined. Instead, they fragment into collimated sprays of hadrons called "jets."
- In many events, jets are produced in pairs. They are called dijets.
- These pairs has defined transverse momentum components which are uniformly distributed in all directions.
- The longitudinal component of momentum along jet is generally not known.



- The dataset is the background dijet sample from the LHC Olympics anomaly detection challenge
- The data is 4 dimensional and is presented as transverse momenta of dijet – 2 back to back jet streams.
- The leading jet is the one with higher transverse momenta even though the transverse momenta of both jets are approximately same.
- Each event is given by $X = (p_{1x}, p_{1y}, p_{2x}, p_{2y})$ where p1 refers to the momentum of the leading jet, p2 represents the momentum of the subleading jet, and x and y are the Cartesian coordinates in the transverse plane



Relative beam sizes around IP1 (Atlas) in collision

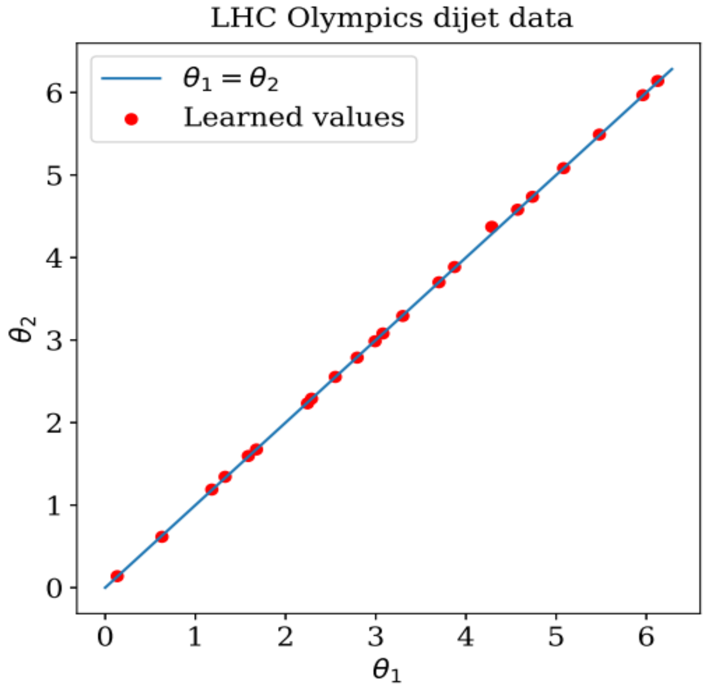
SO(2) X SO(2) searchspace

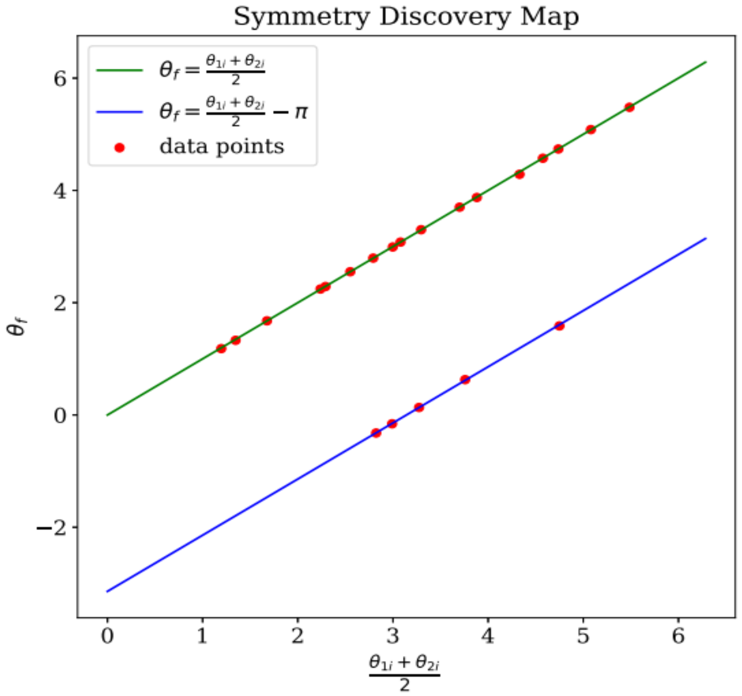
- Both the jets are independently rotated.
- By conservation of momentum , we expect that only those rotations that simultaneously rotate both jets by the same angle will be symmetries.
- To verify our hypothesis, lets train our model parametrisized as

$$g_{\theta_1,\theta_2}\begin{bmatrix} p_{1x} \\ p_{1y} \\ p_{2x} \\ p_{2y} \end{bmatrix} = \begin{bmatrix} \cos\theta_1 & \sin\theta_1 & 0 & 0 \\ -\sin\theta_1 & \cos\theta_1 & 0 & 0 \\ 0 & 0 & \cos\theta_2 & \sin\theta_2 \\ 0 & 0 & -\sin\theta_2 & \cos\theta_2 \end{bmatrix} \begin{bmatrix} p_{1x} \\ p_{1y} \\ p_{2x} \\ p_{2y} \end{bmatrix}$$

- The results are summarized in the graph. SymmetryGAN proved the hypothesis and established the rotational symmetry in SO(2) X SO(2) subspace
- A pattern is observed with the initialized and learned parameters SymmetryGAN. This helps us to map the initial and final parameters and this map is called the Symmetry discovery Map (discussed in detail later). It is given as

$$\Omega(\theta_1,\theta_2) = \begin{cases} \frac{\theta_1+\theta_2}{2} & |\theta_1-\theta_2| < \pi, \\ \frac{\theta_1+\theta_2}{2} - \pi & |\theta_1-\theta_2| > \pi, \end{cases}$$





SO(4) searchspace

- We now turn to the four-dimensional rotation group. SO(4) is a six parameter group, which parametrizes the six independent rotations:

$$R_1 : p_{1x} \rightsquigarrow p_{1y}, \qquad R_2 : p_{1x} \rightsquigarrow p_{2x},$$
$$R_3 : p_{1x} \rightsquigarrow p_{2y}, \qquad R_4 : p_{1y} \rightsquigarrow p_{2x},$$
$$R_5 : p_{1y} \rightsquigarrow p_{2y}, \qquad R_6 : p_{2x} \rightsquigarrow p_{2y},$$

where the notation $R : a \rightsquigarrow b$ means

$$R(a) = a \cos\theta + b \sin\theta,$$
$$R(b) = b \cos\theta - a \sin\theta.$$

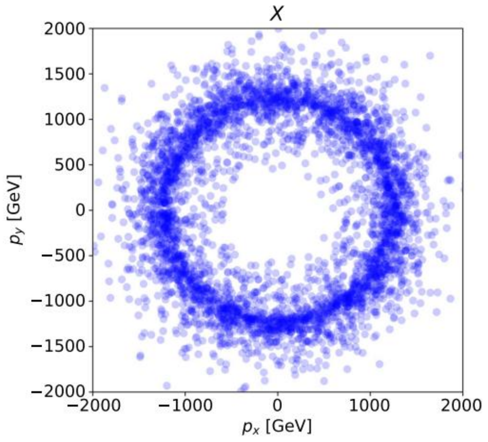
- Such rotations applied back to back on a space constitute the parametrized transformation given by

$$g_{\boldsymbol{\theta}}(X) = R_1 R_2 R_3 R_4 R_5 R_6 X.$$

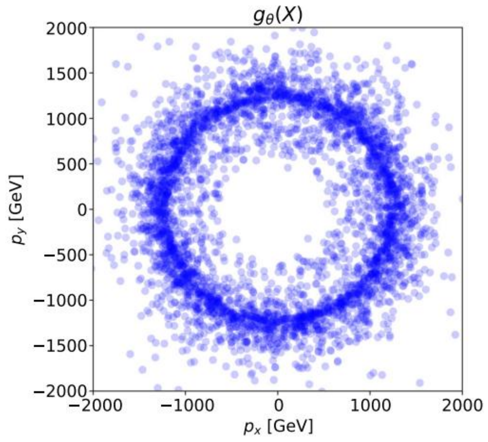
- Since it's a 6 dimensional space, it is hard to visualize the loss landscape to infer the symmetry and to confirm if the learned parameters are actually the symmetries
- There arises a need to verify symmetry leaned by the model.

Symmetry verification strategies

1. To visually inspect the spectra of X and g(X). The graph on the left suggests that the model learned symmetry.



Original data



Transformed data

2. To make use of the fact that the direction of momenta in transverse plane is uniformly distributed. If the transformed momenta preserves the same, it is inferred to be a symmetry.

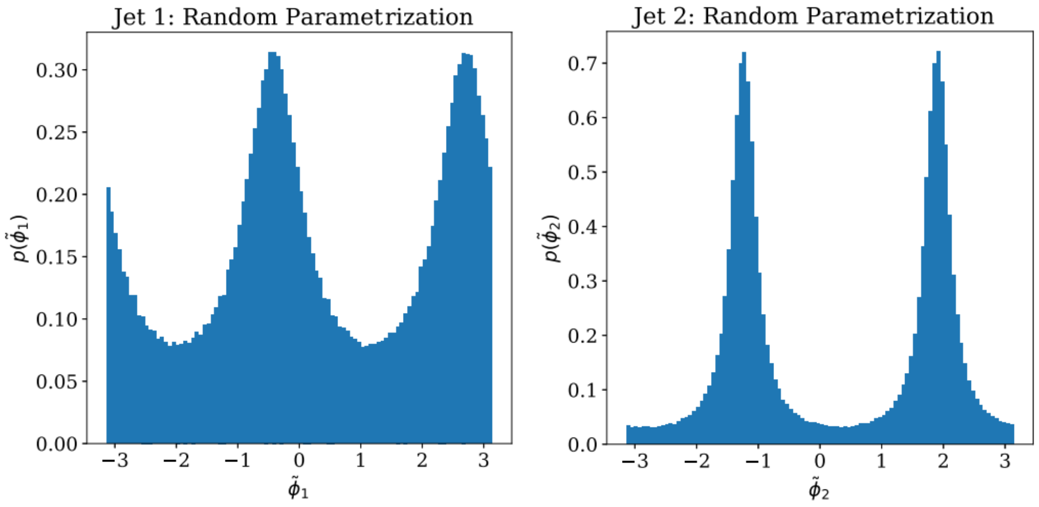
$$X = \begin{bmatrix} p_{1x} \\ p_{1y} \\ p_{2x} \\ p_{2y} \end{bmatrix} = \begin{bmatrix} p_{1T} & \cos\phi_{i1} \\ p_{1T} & \sin\phi_1 \\ p_{2T} & \cos\phi_2 \\ p_{2T} & \sin\phi_2 \end{bmatrix}, \quad \phi_j \sim \mathcal{U}[-\pi, \pi),$$

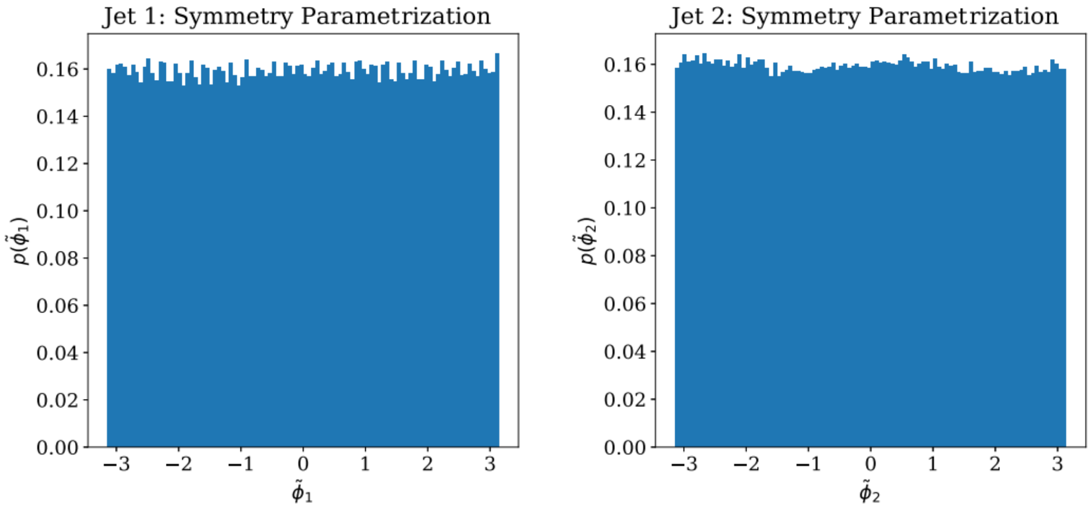
The azimuthal angles of original data is denoted by

$$\phi_j = \arctan 2(p_{jy}, p_{jx})$$

The azimuthal angles of transformed data is denoted by

$$\tilde{\phi}_1 = \arctan 2(g_{\boldsymbol{\theta}}(X)_2, g_{\boldsymbol{\theta}}(X)_1),$$



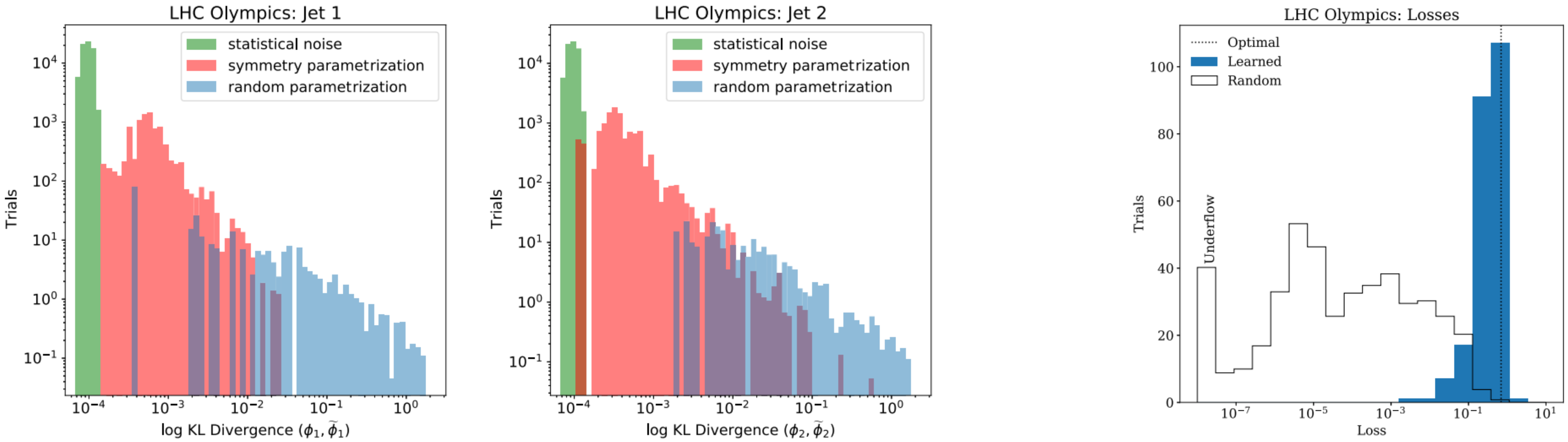


- There is no reason for a non- symmetrical transformation to have uniform distribution of azimuthal angles in U[-π,π) and hence it has a different distribution as seen in the graph
- The difference in the distributions can be quantified by computing the KullbackLeibler (KL) divergence of the two ϕ˜ j distributions against that of ϕj. It is used in statistics to quantify the confidence in approximation of distributions.KL divergence measures how much one distribution Q diverges from another reference distribution P. If Q=P, the KL divergence is zero, meaning the distributions are identical. Smaller the value, less is the difference between the distributions. It is given as,

$$D_{\mathrm{KL}}(P \parallel Q) = \int_{-\infty}^{\infty} P(x) \log\left(\frac{P(x)}{Q(x)}\right) dx$$

- From the graph, it is seen that the KL divergence of randomly selected elements of SO(4) has means of 0.37 and 0.34 for the leading and sub leading jet, while the KL divergence of symmetries in SO(4) has respective means 0.0058 and 0.0090. The statistical noise sampled from U[-π,π) has a mean of 0.0010.

3. Analysing the loss function value which should be (or closer to) 2log2. The graph shows loss of random rotations in SO(4) compared to the loss of rotations learned by SymmetryGAN, overlaid with the analytic loss of a symmetry, 2 log 2.



-->Thus with 3 different verification strategies, it is concluded that Symmetry GAN learned symmetry in SO(4)

Symmetry inference

Symmetry inference presents a promising extension to this paper and is an area of significant research interest. Despite discovering many maps with SymmetryGAN, it's challenging to infer the exact Lie subgroup of SO(4) in the LHC Olympics data. Symmetry discovery identifies points on the Lie group manifold, but inferring the exact group is complex. Given below are three approaches to symmetry inference.

A. Finding discrete subgroups:

- One way to infer symmetries is to look for discrete cyclic symmetric group by Z_n augmenting the extra loss term in loss function.
- Discrete Symmetry subgroups are easy to infer and gives an idea about the bigger symmetry space

B. Group composition

- It is a powerful strategy which makes use of the closure property of groups.
- By combining these discovered symmetries, we can rapidly increase the number of known points on the manifold. This is because the symmetries form a group, ensuring that their composition also results in a valid symmetry.
- The entire symmetry group can be spanned by a set of 'v' different symmetries $\{r_1, \ldots, r_\nu\}$ v is the product of the representation dimension and the number of connected components.
- The major isssue is that ML models give approximations to the symmetry parameters and hence the error gets compounded with each composition.

Symmetry Discovery Map

- A symmetry discovery map connects the initialized parameters of g to the parameters of the learned function. For a k parameter space, the symmetry discover map Ω,

$$\Omega : \mathbb{R}^k \to \mathbb{R}^k,$$

- Symmetry discovery maps have a greater significance as it is a deformation retract: a specific kind of space map which preserves the topological structure.
- The topology of $GL_n(R)$, however, has been studied for over half a century, and the homotopy and homology groups of several nontrivial subgroups of affine groups have been fully determined . Hence, if the symmetry discovery map were known, one could leverage the full scope of algebraic topology and the known results for the linear groups to understand the topology of the symmetry group.

- It is implemented as a neural network where the network learns the parameters. The challenging part is the formulation of the map .
- Let g(x|c) be the symmetry generator, with the parameters c made explicit. Let Ω(c) be a neural network representing the symmetry discovery map. We data is fed to the neural network and its output is fed into SymmetryGAN as input. It is optimized using the modified loss function in the SymmetryGAN network given by

$$L[\Omega, d] = -\sum_{c \in \{c_a\}} \sum_{x \in \{x_i\}} [\log(d(x)) + \log(1 - d(g(x|\Omega(c))))].$$

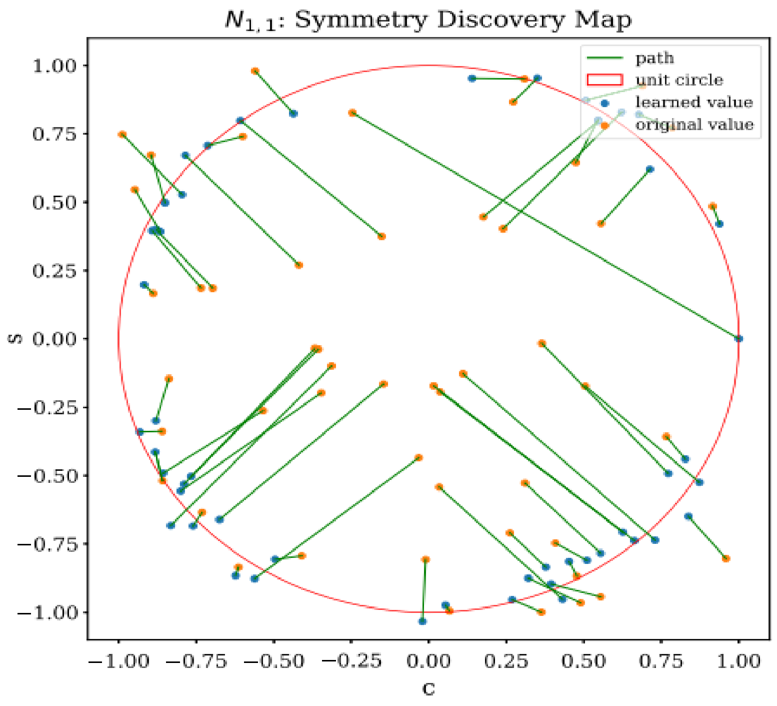
- As long as the symmetry discovery map is sufficiently parametrized and is inizialized as identity, any SymmetryGAN that learns symmetries correctly should also learn the symmetry discovery map correctly.

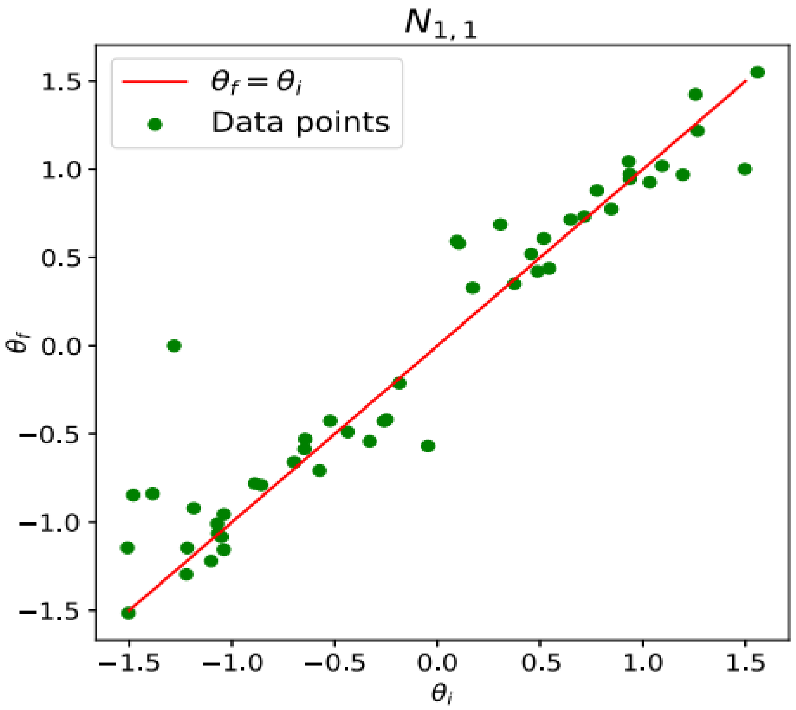
For 1D Gaussian , the map is given as $\Omega(b, c) = \begin{cases} (0, 1) & c > 0, \\ (1, -1) & c < 0. \end{cases}$

For a 2D Gaussian, it is intuitive that the initial parameters will converge in a straight line to its nearest point on the unit circle. This is proven by the graph below. The map is given as below. To make it more visually compelling the map is shown in polar coordinates

$$\Omega(c, s) = \left(\frac{c}{\sqrt{c^2 + s^2}}, \frac{s}{\sqrt{c^2 + s^2}} \right).$$

$$\Omega(r, \theta) = (1, \theta). \qquad r = \sqrt{c^2 + s^2}, \qquad \theta = \arctan 2(s, c).$$





Challenges with Symmetry Discovery Map:

1. For finding the symmetry discovery map, one needs a parametrization of Ω that is flexible enough to describe the map, but simple enough that it can be initialized close to the identity.
2. The Symmetry discovery map is usually discontinuous and this increases the complexity and chances of mode collapse of GANs
3. The nature of activation function makes finding discrete maps challenging. This is explained below

- For a normal distribution N(0,1), g(x)=±x is the symmetry and hence the symmetry discovery map is Ω(c)=sign(c) for g(x)=cx
- To deal with this, a custom activation function like $\Omega(c) = \lambda \mathrm{ReLU}(c) - \mu \mathrm{ReLU}(-c) + \rho.$ can be used. This developes the desired non linearity and can be initialized to unity.
- A more general and robust way is polynomial parametrization ,

$$\Omega(c) = \lambda c + \mu c^2 + \nu c^3 + \cdots + \zeta c^{11},$$

This concludes the presentation