Extended Abstract Track

Continuous Symmetry Discovery and Enforcement for Image Data

Editors: List of editors' names

Abstract

Symmetry is an often-desired quality of machine learning models, leading, among other things, to more predictable model generalization. Continuous symmetry detection and enforcement for machine learning are two related topics that have recently been explored using the Lie derivative along vectors fields, which has led to improved computational outcomes. However, though image data is littered with continuous symmetries under which image classifiers are meant to be invariant, the application of the Lie derivative for the detection and enforcement of these symmetries for image data remains underexplored. In this work, we derive vector field infinitesimal generators for various continuous symmetries for image data. We then use these generators to detect and enforce continuous symmetries in image classifiers. We show that these techniques lead to improved outcomes which are not possible using existing methods.

Keywords: Symmetry Enforcement, Symmetry Detection, Computer Vision, Invariance, Geometric AI

1. Introduction

Symmetry-informed machine learning is a field which has continued to accumulate interest, with previous work demonstrating its effectiveness in improving model performance (Lyle et al., 2020; Bergman, 2019; Craven et al., 2022; Tahmasebi and Jegelka, 2023; Ko et al., 2024). Within the context of image classification tasks, one is often concerned with learning neural networks which are invariant to pre-specified transformations of images. Although a common approach to building invariant models in practice is to explicitly augment the training data with transformed copies of images, recent work has been conducted with the goal of enforcing either equivariance or invariance in models without the need for data augmentation (Finzi et al., 2020). Additionally, recent work has shown that the Lie derivative along vector fields can be used to enforce and discover continuous symmetries in models more generally, which approach can be computationally efficient (Otto et al., 2024; Shaw et al., 2024, 2025). However, it appears that, despite the many continuous symmetries for image data, previous work with regard to image data deals only with certain types of continuous transformations, such as planar affine transformations. In this work, we propose an approach to symmetry discovery and enforcement for image data using the Lie derivative along vector fields. The extension to image data is non-trivial, due to the fact that previous work assumes data is given as a vector in $\mathbb{R}^n$: while an image can be "flattened" so as to be given as a vector, this erases positional structure within images that is commonly leveraged in modern architectures, including Convolutional Neural Networks (CNNs).

2. Related Work

Some related work in symmetry discovery has focused on detecting affine transformation symmetries and encoding the discovered symmetries automatically into a model architecture. Some methods identify Lie algebra generators to describe the symmetries. For example, *augerino* (Benton et al., 2020) attempts to learn a distribution over augmentations, subsequently training a model with augmented data. *Lconv* (Dehmamy et al., 2021), which generalizes CNNs in the presence of affine symmetries, uses infinitesimal generators represented as vector fields to describe the symmetry. SymmetryGAN (Desai et al., 2022) has also been used to detect rotational symmetry (Yang et al., 2023).

Another notable contribution to efforts to detect symmetries of data is *LieGAN*. LieGAN is a generative-adversarial network intended to return infinitesimal generators of the continuous symmetry group of a given dataset (Yang et al., 2023). LieGAN has been shown to detect continuous affine symmetries, including transformations from the Lorentz group. It has also been shown to identify discrete symmetries such as rotations by a fixed angle.

While most continuous symmetry detection methods attempt to discover symmetries which are affine transformations, the representation of infinitesimal generators using vector fields has led to the discovery of continuous symmetries which are not affine (Ko et al., 2024; Shaw et al., 2024). In one case, the domains of image data and partial differential equations are examined in particular (Ko et al., 2024). Compared with other methods, a vector field approach is computationally efficient (Hu et al., 2025).

Our method of symmetry enforcement is most closely related to previous work on symmetry enforcement using vector fields (Finzi et al., 2020; Otto et al., 2024; Shaw et al., 2025), since we are directly adapting the analogous methods presented by the various authors for image data and common image transformations. This method is also analogous to Physics-Informed Neural Networks (PINNs) (Raissi et al., 2019). With PINNs, model training is regularized using differential constraints which represent the governing equations for a physical system. The method of symmetry enforcement employed here differs from PINNs since the differential constraints obtained using infinitesimal generators do not generally have the interpretation of defining governing equations for a physical system.

Continuous symmetry enforcement in images is far from new, and we note some recently-developed methods. Some methods seek to enforce symmetry by augmenting the training dataset according to known symmetries (Bergman, 2019). *Augerino* attempts to enforce symmetry using augmented data, though the symmetries are discovered from the data rather than given *a priori*. Another established method of enforcing symmetry is feature averaging, which is thought to be generally more effective than data augmentation (Lyle et al., 2020).

A growing number of methods seek to use symmetry to construct invariant or equivariant models without the need for augmented training data, and there is previous work accomplishing this using infinitesimal generators (Dehmamy et al., 2021; Yang et al., 2023; Finzi et al., 2020; Otto et al., 2024). Some previous work addresses specific cases: the special case of compact groups is studied (Bloem-Reddy and Teh, 2020), and the case of equivariant CNNs on Homegeneous spaces is studied (Cohen et al., 2019). Other work speaks to the universality of invariant architectures (Maron et al., 2019; Keriven and Peyré, 2019; Yarotsky, 2018).

3. Methods

3.1. Quantifying the Extent to which a Smooth Model is Invariant

The notion of similarity of vector fields has been previously discussed (Shaw et al., 2025), though we include this discussion herein since the proposed methodology is not likely known to the machine learning community at large. Given a metric tensor g_{ij}–usually assumed to be the standard Euclidean metric tensor–we define the angle between two vector fields X and $\hat{X}$ by

$$\cos\left(\theta(X,\hat{X})\right) = \frac{1}{\int_\Omega d\mathcal{M}}\mathbb{E}\left[\frac{|\langle X,\hat{X}\rangle_g|}{||X||_g \cdot ||\hat{X}||_g}\right], \tag{1}$$

where $\langle X,\hat{X}\rangle_g = \sum_{i,j} f_i \hat{f}_j g_{ij}$, $||X||_g = \sqrt{\langle X, X\rangle_g}$, and where $\mathbb{E}\left[u(\mathbf{x})\right] = \int_\Omega u(\mathbf{x}) d\mathcal{M}$, with the region Ω being defined by the range of a given dataset. Ordinarily, this is the full range of the dataset. This formula is a generalization of the formula given in Shaw et al. (2024) in the case where the manifold and/or metric is not assumed to be Euclidean.

The notion of a "cosine similarity" between vector fields induces a method by which the extent to which a particular smooth function is invariant can be quantified in relation to other functions. For a fixed vector field X, a function f is X-invariant if and only if X is orthogonal to the gradient of f, which vector field we denote X_f. Thus, the extent to which f is X-invariant can be quantified in terms of the cosine of the angle between X and X_f, given in Equation (1): the closer to 0 this value is, the more X-invariant f is.

3.2. Continuous Symmetry Enforcement using Vector Field Regularization

In this section, we summarize the previously-given method Shaw et al. (2025); Otto et al. (2024); Finzi et al. (2020) of enforcing continuous symmetries using regularization induced by vector fields. We seek to extend these approaches to image data.

Suppose that, as in the case of supervised learning, one seeks to learn a function F mapping data instances x_i to targets y_i, where $x_i \in \mathbb{R}^d$ and $y_i \in \mathbb{R}^m$. Suppose also that the function F is estimated by means of the minimization of $\mathcal{L}(F(\mathbf{x}),\mathbf{y})$ for a smooth loss function $\mathcal{L}$. The model function F is invariant with respect to the infinitesimal generators $\{X_k\}_{k=1}^s$ precisely when, for each component f_j of F, $X_k(f_j) = 0$ for $1 \leq k \leq s$ and $1 \leq j \leq m$. This can be used as a regularization term giving a loss function of:

$$(1-\lambda(t))\mathcal{L}(F(\mathbf{x}),\mathbf{y}) + \lambda(t)\tilde{\mathcal{L}}(\mathbf{X}(F)(\mathbf{x}),\mathcal{O}), \tag{2}$$

where $\mathbf{X}(F)(\mathbf{x}) = (X_k(f_j))(x_i)$, $\mathcal{O}$ is an array of zeros and is of the same shape as $\mathbf{X}(F)(\mathbf{x})$, $\tilde{\mathcal{L}}$ is a smooth loss function, and $\lambda(t) \in [0,1]$ is a (possibly time/epoch-dependent) symmetry regularization parameter.

4. Experimental Results

We have performed experiments using the MNIST dataset (Deng, 2012), comparing median $\pm$ IQR/2 accuracy scores using augmentation only, symmetry regularization, regularization and augmentation, and a baseline with neither augmentation nor regularization. Table 1 shows the results of our experiment using Power Law symmetry, and Table 2 shows the results using Gaussian Blur.

Table 1: Power Law Symmetry Enforcement on the MNIST Dataset

γ	**Baseline**	**Regularization**	**Augmentation**	**Reg + Aug**
0.1	0.4435 ± 0.0757	0.7289 ± 0.0528	0.9643 ± 0.0318	0.9066 ± 0.0438
0.2	0.8670 ± 0.0332	0.9497 ± 0.0365	0.9757 ± 0.0315	0.9275 ± 0.0446
0.25	0.9309 ± 0.0359	0.9604 ± 0.0357	0.9765 ± 0.0317	0.9295 ± 0.0446
1.0	0.9684 ± 0.0413	0.9690 ± 0.0355	0.9773 ± 0.0323	0.9317 ± 0.0449

For the Power Law experiment, each model was trained for a total of 12 epochs. For the Regularization and Reg + Aug methods, $\lambda(t)$ was 0 for the first three epochs and was subsequently 0.5. With further regard to Equation (2), the Cross Entropy loss was used for $\mathcal{L}$, while the MSE loss multiplied by 28^2 was used for $\tilde{\mathcal{L}}$. Optimization was obtained using the Adam optimizer with a learning rate of 0.001. For the Gaussian blur experiment, nearly identical settings were used, except that $\tilde{\mathcal{L}}$ was scaled by a factor of $10 \cdot 28^2$ rather than 28^2, as before.

Table 2: Gaussian Blur Symmetry Enforcement on the MNIST Dataset

σ	**Baseline**	**Regularization**	**Augmentation**	**Reg + Aug**
0.1	0.8924 ± 0.0024	0.9710 ± 0.0439	0.9216 ± 0.0409	0.9797 ± 0.0022
1.0	0.8867 ± 0.0027	0.9594 ± 0.0421	0.9225 ± 0.0405	0.9714 ± 0.0034
2.0	0.8451 ± 0.0138	0.8731 ± 0.0438	0.9167 ± 0.0391	0.8969 ± 0.0022
3.0	0.7747 ± 0.0201	0.7524 ± 0.0618	0.9103 ± 0.0371	0.8108 ± 0.0156
6.0	0.7076 ± 0.0354	0.6288 ± 0.0389	0.8998 ± 0.0346	0.7234 ± 0.0485

5. Conclusion

We have introduced symmetry discovery and enforcement for image classifiers using vector fields. We have shown that symmetry enforcement can encourage models to be more symmetric than a baseline approach. Future work includes the examination of the effect of the regularization parameter $\lambda(t)$ on model performance and generalization, as well as incorporating additional symmetry types and image datasets in experiments.

References

Gregory Benton, Marc Finzi, Pavel Izmailov, and Andrew G Wilson. Learning invariances in neural networks from training data. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 17605–17616. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/cc8090c4d2791cdd9cd2cb3c24296190-Paper.pdf.

Doron L. Bergman. Symmetry constrained machine learning. In *Advances in Intelligent Systems and Computing*, pages 501–512. Springer International Publishing, aug 2019. doi: 10.1007/978-3-030-29513-4_37. URL https://doi.org/10.1007.

Benjamin Bloem-Reddy and Yee Whye Teh. Probabilistic symmetries and invariant neural networks. *J. Mach. Learn. Res.*, 21(1), January 2020. ISSN 1532-4435.

Taco S. Cohen, Mario Geiger, and Maurice Weiler. A general theory of equivariant cnns on homogeneous spaces. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, Red Hook, NY, USA, 2019. Curran Associates Inc.

Sean Craven, Djuna Croon, Daniel Cutting, and Rachel Houtz. Machine learning a manifold. *Physical Review D*, 105(9), may 2022. doi: 10.1103/physrevd.105.096030. URL https://doi.org/10.1103.

N. Dehmamy, R. Walters, Y. Liu, D. Wang, and R. Yu. Automatic symmetry discovery with lie algebra convolutional network. *Advances in Neural Information Processing Systems*, 34:2503–2515, 2021. URL https://arxiv.org/abs/2109.07103.

Li Deng. The mnist database of handwritten digit images for machine learning research [best of the web]. *IEEE Signal Processing Magazine*, 29(6):141–142, 2012. doi: 10.1109/MSP.2012.2211477.

Krish Desai, Benjamin Nachman, and Jesse Thaler. Symmetry discovery with deep learning. *Phys. Rev. D*, 105:096031, May 2022. doi: 10.1103/PhysRevD.105.096031. URL https://link.aps.org/doi/10.1103/PhysRevD.105.096031.

Marc Finzi, Samuel Stanton, Pavel Izmailov, and Andrew Gordon Wilson. Generalizing convolutional neural networks for equivariance to lie groups on arbitrary continuous data. In *Proceedings of the 37th International Conference on Machine Learning*, ICML'20. JMLR.org, 2020.

Lexiang Hu, Yikang Li, and Zhouchen Lin. Explicit discovery of nonlinear symmetries from dynamic data. In *Forty-second International Conference on Machine Learning*, 2025.

Nicolas Keriven and Gabriel Peyré. Universal invariant and equivariant graph neural networks. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, Red Hook, NY, USA, 2019. Curran Associates Inc.

Gyeonghoon Ko, Hyunsu Kim, and Juho Lee. Learning infinitesimal generators of continuous symmetries from data. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang, editors, *Advances in Neural Information Processing Systems*, volume 37, pages 85973–86003. Curran Associates, Inc., 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/9c26a48b7ebd37bc3e979c8651fc3649-Paper-Conference.pdf.

John M. Lee. *Introduction to Smooth Manifolds*. Springer New York, NY, 2012.

C. Lyle, M. Wilk, M. Kwiatkowska, Y. Gal, and B. Bloem-Reddy. On the benefits of invariance in neural networks, 2020. URL https://arxiv.org/abs/2005.00178.

Haggai Maron, Ethan Fetaya, Nimrod Segol, and Yaron Lipman. On the universality of invariant networks. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 4363–4371. PMLR, 09–15 Jun 2019. URL https://proceedings.mlr.press/v97/maron19a.html.

Samuel E. Otto, Nicholas Zolman, J. Nathan Kutz, and Steven L. Brunton. A unified framework to enforce, discover, and promote symmetry in machine learning, 2024. URL https://arxiv.org/abs/2311.00212.

Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

Ben Shaw, Abram Magner, and Kevin Moon. Symmetry discovery beyond affine transformations. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang, editors, *Advances in Neural Information Processing Systems*, volume 37, pages 112889–112918. Curran Associates, Inc., 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/cd04ec5aebfbe397c7fd718c35d02e0b-Paper-Conference.pdf.

Ben Shaw, Sasidhar Kunapuli, Abram Magner, and Kevin R. Moon. Lie group symmetry discovery and enforcement using vector fields, 2025. URL https://arxiv.org/abs/2505.08219.

Behrooz Tahmasebi and Stefanie Jegelka. The exact sample complexity gain from invariances for kernel regression. *Advances in Neural Information Processing Systems*, 36, 2023.

J. Yang, R. Walters, N. Dehmamy, and R. Yu. Generative adversarial symmetry discovery. *Proceedings of the 40th International Conference on Machine Learning*, 202(1647):39488–39508, 2023. URL https://arxiv.org/abs/2302.00236.

Dmitry Yarotsky. Universal approximations of invariant maps by neural networks. *Constructive Approximation*, 55:407–474, 2018. URL https://api.semanticscholar.org/CorpusID:13745401.

Appendix A. Vector Fields and Flows

We now provide some background on vector fields and their associated flows (1-parameter transformations). We refer the reader to literature on the subject for additional information Lee (2012). Suppose that X is a smooth (tangent) vector field on $\mathbb{R}^n$:

$$X = \alpha^i \partial_{x^i} := \sum_{i=1}^{n} \alpha^i \partial_{x^i}, \tag{3}$$

where $\alpha^i : \mathbb{R}^n \to \mathbb{R}$ for $i \in [1, n]$, and where $\{x^i\}_{i=1}^n$ are coordinates on $\mathbb{R}^n$. X assigns a tangent vector at each point and can also be viewed as a function on the set of smooth, real-valued functions. E.g. if $f : \mathbb{R}^n \to \mathbb{R}$ is smooth,

$$X(f) = \sum_{i=1}^{n} \alpha^i \frac{\partial f}{\partial x^i}. \tag{4}$$

For example, for $n = 2$, if $f(x, y) = xy$ and $X = y\partial_x$, then $X(f) = y^2$. X is also a *derivation* on the set of smooth functions on $\mathbb{R}^n$: that is, for smooth functions $f_1, f_2 : \mathbb{R}^n \to \mathbb{R}$ and $a_1, a_2 \in \mathbb{R}$,

$$X(a_1 f_1 + a_2 f_2) = a_1 X(f_1) + a_2 X(f_2), \qquad X(f_1 f_2) = X(f_1) f_2 + f_1 X(f_2). \tag{5}$$

These properties are satisfied by derivatives. A flow on $\mathbb{R}^n$ is a smooth function $\Psi : \mathbb{R}\times\mathbb{R}^n \to \mathbb{R}^n$ which satisfies

$$\Psi(0, p) = p, \qquad \Psi(s, \Psi(t, p)) = \Psi(s + t, p) \tag{6}$$

for all $s, t \in \mathbb{R}$ and for all $p \in \mathbb{R}^n$. A flow is a 1-parameter group of transformations. An example of a flow $\Psi : \mathbb{R} \times \mathbb{R}^2 \to \mathbb{R}^2$ is

$$\Psi(t, (x, y)) = (x \cos(t) - y \sin(t), x \sin(t) + y \cos(t)), \tag{7}$$

with t being the continuous parameter known as the flow parameter. This flow rotates a point (x, y) about the origin by t radians.

For a given flow Ψ, one may define a (unique) vector field X as given in Equation 4, where each function α^i is defined as

$$\alpha^i = \left(\frac{\partial \Psi}{\partial t}\right)\bigg|_{t=0}. \tag{8}$$

Such a vector field is called the infinitesimal generator of the flow Ψ. For example, the infinitesimal generator of the flow given in Equation 7 is $-y\partial_x + x\partial_y$.

Conversely, given a vector field X as in Equation 4, one may define a corresponding flow as follows. Consider the following system of differential equations:

$$\frac{dx^i}{dt} = \alpha^i, \qquad x^i(0) = x^i_0. \tag{9}$$

Suppose that a solution $\mathbf{x}(t)$ to Equation 9 exists for all $t \in \mathbb{R}$ and for all initial conditions $\mathbf{x}_0 \in \mathbb{R}^n$. Then the function $\Psi : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ given by

$$\Psi(t, \mathbf{x}_0) = \mathbf{x}(t) \tag{10}$$

is a flow. The infinitesimal generator corresponding to Ψ is X. For example, to calculate the flow of $-y\partial_x + x\partial_y$, we solve

$$\dot{x} = -y, \quad \dot{y} = x, \qquad x(0) = x_0, \quad y(0) = y_0 \tag{11}$$

and obtain the flow $\Psi(t, (x_0, y_0))$ defined by Equation 7. It is generally easier to obtain the infinitesimal generator of a flow than to obtain the flow of an infinitesimal generator.

A smooth function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be X-invariant if $X(f) = 0$ identically for a smooth vector field X. The function f is Ψ-invariant if, for all $t \in \mathbb{R}$, $f = f(\Psi(t, \cdot))$ for a flow Ψ. If X is the infinitesimal generator of Ψ, f is Ψ-invariant if and only if f is X-invariant.

Appendix B. Infinitesimal Generators of Multi-Parameter Groups

Let $G \in \mathbb{R}^s$ be a group, and suppose G acts on $\mathbb{R}^n$: that is, for $g_1, g_2 \in G$ and for $x \in \mathbb{R}^n$, there is a function $\Psi : G \times \mathbb{R}^n \to \mathbb{R}^n$ such that (assuming the group operation is vector addition)

$$\Psi(\mathbf{0}, x) = x, \qquad \Psi(g_2, \Psi(g_1, x)) = \Psi(g_1 + g_2, x). \tag{12}$$

The use of the symbol Ψ to denote a multi-parameter group action is not accident, as a flow is a 1-parameter group action. Let $\{v_i\}_{i=1}^s$ be a basis for the tangent space of G at $\mathbf{0}$, the group identity element. Lastly, let σ be a curve in G for which $\sigma(t_0) = \mathbf{0}$ and $\dot{\sigma}(t_0) = v_i$ for $t_0 \in \mathbb{R}$. The infinitesimal generator X_i corresponding to v_i is given by

$$X_i = \left(\frac{d}{dt} \Psi(\sigma(t), x) \right) \bigg|_{t=t_0}. \tag{13}$$

For example, consider the group $G = \mathbb{R}^3$ acting on $\mathbb{R}^2$ via

$$\Psi((a, b, \theta), (x, y)) = (x \cos(\theta) - y \sin(\theta) + a, x \sin(\theta) + y \cos(\theta) + b).$$

Given the following three curves,

$$\sigma_a(t) = (t, 0, 0), \qquad \sigma_b(t) = (0, t, 0), \qquad \sigma_\theta(t) = (0, 0, t),$$

we find that

$$X_a = \frac{d}{dt}(t, 0)\,|_{t=0} = \partial_x, \qquad X_b = \frac{d}{dt}(0, t)\,|_{t=0} = \partial_y,$$

$$X_\theta = \frac{d}{dt}(x \cos(\theta) - y \sin(\theta), x \sin(\theta) + y \cos(\theta))\,|_{t=0} = -y\partial_x + x\partial_y.$$

For each of these vector fields, a corresponding flow can be computed, which flows we call Ψ_a, Ψ_b, and Ψ_θ, respectively. In terms of the original parameters, these flows are given as

$$\Psi_a(a, (x, y)) = (x + a, y), \qquad \Psi_b(b, (x, y)) = (x, y + b),$$

$$\Psi_\theta = (x \cos(\theta) - y \sin(\theta), x \sin(\theta) + y \cos(\theta)).$$

While each of these flows are, individually, 1-parameter group actions, it is clear that the infinitesimal generators X_a, X_b, and X_θ are the infinitesimal generators for the multi-parameter group action given in Equation (12). Thus, discovering vector field infinitesimal generators which annihilate a fixed (smooth) function applies to multi-parameter group actions and not solely to 1-parameter groups.

Appendix C. Diagonal Group Actions and their Infinitesimal Generators

Let G be a group acting on a set Y. The action of G on Y induces an action on the set Y^m, where a group element $g \in G$ acts on $(y_1, y_2, \ldots, y_m) \in Y^m$ by acting on each component y_i separately and in the manner in which g acts on $y \in Y$. It is said that the action of G on Y^m is a diagonal action.

If $\Psi(t, \mathbf{x})$ is a flow on $\mathbb{R}^n$, the induced diagonal action on $\mathbb{R}^n \times \mathbb{R}^n \times \cdots \times \mathbb{R}^n = (\mathbb{R}^n)^m$ is defined as

$$\tilde{\Psi}(t, (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m)) = (\Psi(t, \mathbf{x}_1), \Psi(t, \mathbf{x}_2), \ldots, \Psi(t, \mathbf{x}_m)) .$$

If X is the vector field infinitesimal generator for $\Psi(t, \mathbf{x})$, we denote the infinitesimal generator for $\Psi(t, \mathbf{x}_i)$ by X_i, so that the infinitesimal generator $\tilde{X}$ for $\tilde{\Psi}$ is given as

$$\tilde{X} = X_1 + X_2 + \cdots + X_m.$$

For example, let $\Psi(t, (x, y)) = (x + t, ty)$, so that $X = \partial_x + y\partial_y$, the level curves for which can be written as $y = ce^x$. The infinitesimal generator for the diagonal action of Ψ on $\mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^2$ is

$$\tilde{X} = \partial_{x_1} + y_1\partial_{y_1} + \partial_{x_2} + y_2\partial_{y_2} + \partial_{x_3} + y_3\partial_{y_3}.$$

Our model assumption is that an image with m channels is an element of $(\mathbb{R}^n)^m$, with n being the number of pixels. This merely formalizes the concept of a group "acting on each image channel separately." However, another fruitful perspective outside the context of image data deals with samples, where m represents the number of data points in a given dataset and where n represents the number of features. It is within this context that the sample mean can be said to be *equivariant* under certain affine transformations: a group acts on each point $\{\mathbf{x}_i\}_{i=1}^m$ separately, where $\mathbf{x}_i \in \mathbb{R}^n$, so that the sample mean of $\tilde{\Psi}(t, (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m))$ coincides with $\Psi(t, \cdot)$ applied to the sample mean of $\{\mathbf{x}_i\}_{i=1}^m$.[1]

1. We note that equivariance under flows, like invariance, can be characterized in terms of the Lie derivative (Otto et al., 2024), though the subject of the current work is invariance and not equivariance.