Identifiability of Discretized Latent Coordinate Systems via Density Landmarks Detection

Vitória Barin-Pacela [1 2] **Kartik Ahuja** [1] **Simon Lacoste-Julien** [2 3] **Pascal Vincent** [1 2 4]

Abstract

Disentanglement aims to recover meaningful latent ground-truth factors from only the observed distribution. Identifiability provides the theoretical grounding for disentanglement to be well-founded. Unfortunately, unsupervised identifiability of independent latent factors is a theoretically proven impossibility in the i.i.d. setting under a general nonlinear smooth map from factors to observations. In this work, we show that, remarkably, it is possible to recover discretized latent coordinates under a highly generic nonlinear smooth mapping (a diffeomorphism) without any additional inductive bias on the mapping. This is, assuming that latent density has axis-aligned discontinuity landmarks, but without making the unrealistic assumption of statistical independence of the factors. We introduce this novel form of identifiability, termed *quantized coordinate identifiability*, and provide a comprehensive proof of the recovery of discretized coordinates.

1. Introduction

A large part of intelligence is based on the ability to make sense of observed sensory data, without necessarily explicit supervision. The goal of representation learning is, thus, to detect and model relevant structure in the distribution of observed data, and expose it into useful compact representations, to facilitate good generalization and sample-efficient learning of subsequent tasks. One long-standing goal in that respect has been that of structuring the representation into *disentangled factors* (Bengio et al., 2013). These may be conceived of as "natural" ground truth, descriptive, or causal variables that underlie the observations. A vector representation that is made of recovered disentangled factors may be viewed as corresponding to a natural Cartesian coordinate system for the observations, whereby each varying factor is associated with an axis.

Identifiability theory formalizes the foundations of disentanglement by precisely delimiting conditions under which it is possible. Unsupervised identifiability of independent latent factors has been proven impossible in the general nonlinear setting, in the absence of further inductive bias (Hyvärinen & Pajunen, 1999; Locatello et al., 2019). In the present theoretic work, we revisit and tackle the problem of fully unsupervised identifiability of latent factors in the same challenging setting: the most general form of smooth non-linear mapping, a diffeomorphism. No additional assumptions are made on the mapping, and the assumption of the factors being mutually independent is also removed.

We replace the latter by assuming the presence of axis-aligned discontinuities in the joint probability density of the latent factors; such discontinuities uncover sufficient information to enable a relaxed form of identifiability of the latent factors. This is a remarkable result, given the significant distortion of space caused by a diffeomorphism. First, we show that the discontinuities are preserved under diffeomorphisms. These discontinuities, or cliffs, can act as "separators" of the density into different regions. We prove that under these assumptions, it is possible to detect these separators that allow us to partition input space. In this way, we can recover a Cartesian coordinate system yielding *discretized* coordinates of the latent factors. Therefore, we are able to map observed points back into their respective factor bins. This theoretical work aims to lay out the foundations for this relaxed form of identifiability, in hopes of leading to algorithms of practical relevance, since it requires neither the restrictive assumptions on the mapping's function class nor the unrealistic assumption of independence to achieve identifiability.

Existing work in the literature achieves identifiability by making assumptions a) on both the mixing map and the latent factors, which are typically fully unsupervised; b) assumptions on the distribution of latent factors and not

[1]FAIR, Meta AI, Montréal, Canada [2]Mila and DIRO, Université de Montréal, Montréal, Canada [3]Canada CIFAR AI Chair [4]CIFAR, Canada. Correspondence to: Vitória Barin-Pacela <vitoria.barin-pacela@mila.quebec>.

Accepted to the LXAI Workshop at ICML, Honolulu, Hawaii, USA.
Copyright 2023 by the author(s).

strictly on the mixing map, which typically require weak supervision or auxiliary variables. None of these studies considered the recovery of a discretized coordinate system like we do in this work.

Identifiability of latent factors in the unsupervised i.i.d setting: The seminal work on linear independent component analysis (Comon, 1994) established that under a linear and invertible mixing map and independent non-Gaussian latent factors, we can identify these latent factors up to order and scale indeterminacies. Taleb & Jutten (1999) restrict the problem to a post-nonlinear mapping, obtaining the same indeterminacies as in linear mixtures. Gresele et al. (2021) demonstrated that with independent latent factors and a mixing function that adheres to the independent mechanism assumption, some of the non-identifiability counterexamples highlighted in (Hyvärinen & Pajunen, 1999) can be avoided. Expanding on the role of mixing maps, Buchholz et al. (2022) scrutinized different classes of maps that restrict the Jacobian of the mixing maps. Their study specifically focused on conformal maps and orthogonal coordinate transformations. Lastly, Ahuja et al. (2022) asserted that the true latent factors can be identified, barring permutation and scaling errors, when the mixing map is polynomial and latent factors satisfy the support independence assumption, as proposed in (Wang & Jordan, 2021; Roth et al., 2022). These two aforementioned works relax the assumption of independent factors to that of independent support, leveraging a trivial axis-aligned "grid" structure as the boundary of the support. In the present work, we also leverage axis-aligned structure *inside* the support.

Identifiability of latent factors with weak supervision: Research in this category largely makes assumptions on the latent distribution but imposes few constraints on the mixing map. To compensate for this lack of restrictions, these studies necessitate additional information, typically in one of two forms: a) identification driven by auxiliary information (e.g., labels, time stamps), or b) identification driven by weak supervision (e.g., data augmentations) (Hyvärinen & Morioka, 2017; Hyvärinen et al., 2019; Hyvärinen & Morioka, 2016). A key example of auxiliary information-driven identification is the work on identifiable variational autoencoders (Khemakhem et al., 2020), which assumes the existence of an additionally observed variable such that the latent variables are conditionally independent given it, and the conditional probability density of the latent variable given this auxiliary variable comes from an exponential family.

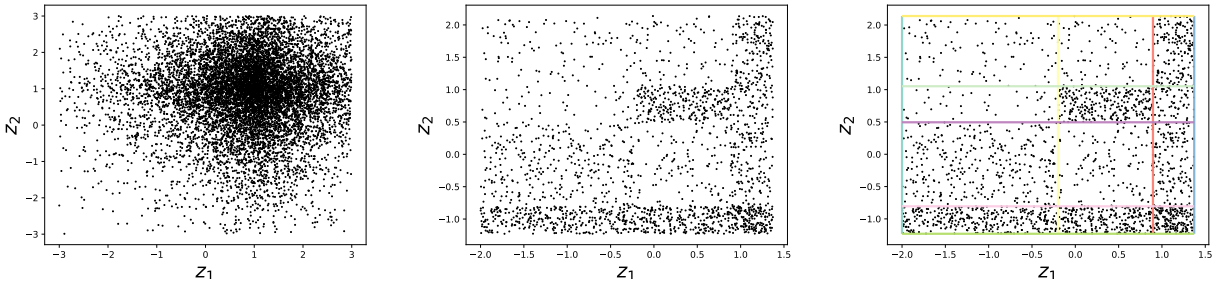
2. Overview of the proposed approach

We suppose that we have access to observations in $\mathcal{X} \subset \mathbb{R}^D$. They are realizations of the vector random variable $X = (X_1, \ldots, X_D)$, which is assumed to be a transformation of a real vector of unobserved latent factors $Z = (Z_1, \ldots, Z_d)$, i.e. $X = f(Z)$, via a *bijective* mapping $f : \mathcal{Z} \to \mathcal{X}$ where $\mathcal{Z} \subset \mathbb{R}^d$. f is called the *mixing map*. f is unknown but is assumed to belong to a broad function class. Latent factors Z follow a distribution represented by probability density function (PDF) p_Z, which is also unknown, but on which assumptions are typically made. This will induce a distribution for X whose PDF is denoted by p_X. The goal of disentanglement is, from observed X only, to recover an inverse mapping $g : \mathcal{X} \to \mathcal{Z}$ that approximates f^{-1} , so that in $Z' = g(X)$, ideally, we have a permutation σ such that each Z'_i correspond to a $Z_{\sigma(i)}$ up to some monotonic transformation. We consider the mixing map to be the most general form of smooth invertible mapping: f is a *diffeomorphism*, that is, a continously differentiable function with continously differentiable inverse. We want to learn the approximate inverse diffeomorphism g. Most of the theory will concern the diffeomorphism $h := g \circ f$ that maps the two latent spaces together. The setup is summarized in the following diagram:

$$\underbrace{Z}_{\substack{\subset \mathbb{R}^d \\ \text{true factors}}} \sim p_Z \xrightarrow[\text{unknown}]{f} \underbrace{X}_{\substack{\subset \mathbb{R}^D \\ \text{observed data}}} \xrightarrow[\text{learned}]{g \approx f^{-1}} \underbrace{Z'}_{\substack{\subset \mathbb{R}^d \\ \text{recovered factors}}}, \qquad h = g \circ f$$

2.1. Principle of our approach

Statistical independence of latent factors has been criticized as an unrealistic and problematic assumption (Träuble et al., 2021; Dittadi et al., 2021; Roth et al., 2022) whose association to disentanglement is misleading. In our proposed approach, we do not assume that the factors Z_i's are independent of each other. Instead, we assume that the probability density landscape of p_Z has *remarkable landmarks*, such as cliffs, that are expected to be *axis-aligned* in the ground-truth latent space. Such landmarks need to be detectable not just in p_Z but also in p_X. This means that they must correspond to events that are sufficiently striking and that the mixing map f is correspondingly gentle enough that it can neither erase nor create them. Since we consider the most general non-linear smooth maps – *diffeomorphisms* – which can warp the space in almost arbitrary ways, the correspondingly striking enough events that can survive these are *discontinuities* in the probability density landscape. Note that the extreme flexibility of diffeomorphisms is the fundamental reason for the negative non-identifiability results for such general non-linear function class (Locatello et al., 2019). A diffeomorphism can morph almost any continuous distribution into any other continuous distribution, rendering most assumptions on the factors powerless. Even though they are smooth, diffeomorphisms can get infinitesimally close to non-smooth behavior if needed. But one aspect they can neither erase nor manufacture is discontinuities.



(a) Independent non-Gaussian factors

(b) Latent factors whose PDF has axis-aligned discontinuities

(c) Underlying grid structure

Figure 1: Illustration of different kinds of assumptions on the distribution of latent factors. **Left:** samples from traditional assumption of independent non-Gaussian factors (here using a truncated Laplace distribution). **Middle:** samples from a distribution that follows our assumption of axis-aligned discontinuities in the probability density. **Right:** Underlying grid structure revealing discontinuity cliffs in the density landscape as colored *axis-separators*, forming a *grid*. Traditional independence assumption yields non-identifiability result under general nonlinear smooth mapping (diffeomorphism). Our assumption yields, under diffeomorphism, provable recovery of a discretized coordinate system. It allows to map back observed points into the proper latent grid cell – a novel relaxed form of identifiability, which we term *quantized coordinate identifiability*.

We prove the preservation of discontinuities in the PDF under a diffeomorphism (see *Non-removable discontinuity preservation theorem* in Appendix B.1). We can thus hope to detect density landmaks made of discontinuities *because* discontinuities are preserved by the mapping. Our precise assumption is that there are density jumps (discontinuities) in the PDF p_Z of the ground truth factors, characteristic of each factor, so that their *location* does not depend on the values of the other factors. With this assumption, these landmarks form an axis-aligned *grid* on $\mathcal{Z} \subset \mathbb{R}^d$. Figure 1 gives an illustrative 2D example of such a density landscape, contrasting it with an example PDF of statistically independent factors. These discontinuities are preserved by mapping h. Under these circumstances, we can show that it suffices to ensure that g yields an axis-aligned grid in $\mathcal{Z}'$ for us to recover the ground truth coordinate system (up to coordinate permutation) *discretized* to grid-cells. This is our main theoretical contribution in this work: the corresponding identifiability theorem is presented in Section 3. We achieve a relaxed form of identifiability, where in a sense we lose "resolution" on the latent factors, but are still able to disentangle them and bin them into meaningful intervals. With the discretized grid coordinates, we can recover precisely in *which cell* of the grid the factors lie (their discrete coordinates), but not precisely where inside the cell (which would require perfect recovery of the real coordinates). Our identification is actually "up to" arbitrary diffeomorphisms *within* each cell.

Realistic or not realistic? At first sight, it may seem that we are getting rid of an unrealistic assumption – that of statistical independence of factors – to replace it with another one that is on the surface seemingly even less realistic: grid-forming discontinuities in the probability density landscape. But is that so? It depends if one thinks of disentangled factors from the ingrained viewpoint of independence, or form the more useful perspectives of either descriptive explanatory factors or factors that have a causal origin. The usual concrete *descriptive factors* with which we tend to describe things are clearly not statistically independent (color of a banana v.s. orange-shaped object; sand v.s. grass background for the picture of a camel v.s. cow). On the other hand, examples of fundamental sharp changes in the probability density along a given factor abound. Due to gravity, people and most objects tend to be either in a standing or lying position. One will seldom see them with a 45 degrees pitch angle irrespective of how other factors appear, which gives rise to the axis-aligned nature of these sharp changes in density. While sharp density changes are not necessarily discontinuities, they may be conceived of as smoothed discontinuities.

For example, we find evidence of grid structure in the real-world dataset of exoplanets, shown in Figure 2. The NASA Exoplanet Archive (Akeson et al., 2013) contains physical measurements of exoplanets. Sharp density changes are observed, e.g. in the descriptive factors *stellar magnitude* and *radius multiplier*. Interestingly, the magnitude of the joint probability density gradient of these variables displays cliffs that appear axis-aligned, hinting at a grid structure that resembles the synthetic data from Figure 1, thus substantiating the realism of this assumption.

Note also that while the picture of a grid with many

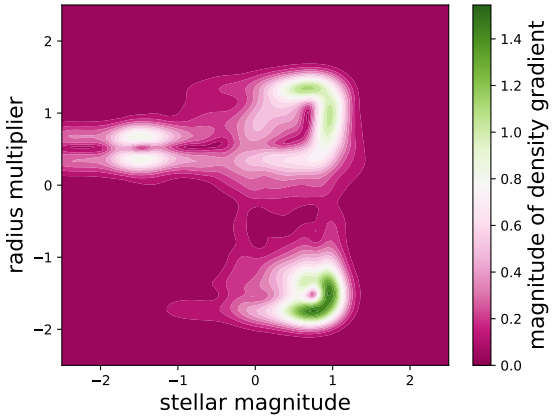


Figure 2: Grid structure observed in an exoplanet dataset.

separators per axis may look unrealistic, finding merely one or two such separator landmarks per axis will already provide a precise location among 2^K or 3^K cells, K being the number of factors.

While our work is agnostic to how such structures in the density might emerge, they may be motivated from a causal perspective, i.e. if the ground truth factors are associated with latent *causal factors*. From this perspective, what we are uncovering in aiming for such cliff-like density landmarks are *causal footprints* (Lopez-Paz et al., 2017).

Finally, we highlight that beyond the presence of axis-aligned discontinuities, we make no assumptions at all on what the factor's distribution should be like outside of these localized discontinuities.

3. Discretized coordinates identifiability

We here briefly present the main theoretical identifiability result of this study (corresponds to Theorem 4, formalized in Appendix B.3, with a complete proof). We use the following definitions:

An *axis-separator* is a an axis-aligned hyperplane (coordinate hyperplane) restricted to the support of the density (as the colored segments in Figure 1c).

An *axis-aligned grid* is a union of axis-separators.

A *discrete coordination* $\mathbf{A}$ represents the coordinates of axis-separators along each of the coordinate axes, defining the entire grid structure. For instance $\mathbf{A}_2 = (-5.2, 0.7, 2.6)$ means that there are 3 axis-separators on the second axis, defined respectively as level sets $z_2 = -5.2$; $z_2 = 0.7$; $z_2 = 2.6$

Discretized coordinates $\bar{z}$ associated to a point with real-valued coordinates z are the integer location of the grid cell that z belongs to, and are obtained as $\bar{z}_i = q(z_i, \mathbf{A}_i, s_i)$, where $\mathbf{A}$ is the discrete coordination, and $s_i \in \{-1, 1\}$ indicate axis reversals[1]. So e.g. if $\mathbf{A}_2 = (-5.2, 0.7, 2.6)$ and $s_2 = +1$ then for $z_2 = 0.9$, we would have $\bar{z}_2 = q(z_2, \mathbf{A}_2, s_2) = 2$.

Discretized coordinates identifiability theorem:

Let Z be a latent random variable with values in $\mathcal{Z} \subset \mathbb{R}^d$ and whose PDF is p_Z. Let $f : \mathcal{Z} \to \mathcal{X} \subset \mathbb{R}^D$ be a diffeomorphism, and $X = f(Z)$ be the observed random variable. Further assume that the PDF p_Z has non-removable discontinuities that forms an axis-aligned grid, whose discrete coordination is $\mathbf{A}$. There exists diffeomorphisms $g : \mathcal{X} \to \mathcal{Z}'$ yielding a variable $Z' = g(X)$ whose PDF $p_{Z'}$ has non-removable discontinuities that form an axis-aligned grid. Consider any such diffeomorphism g, and let $\mathbf{B}$ be the discrete coordination of its resulting axis-aligned grid. Then there exists a permutation function σ over dimension indexes $1, \ldots, d$ and a direction reversal vector $s \in \{-1, +1\}^d$ such that $q(Z'_j, \mathbf{B}_j, 1) = q(Z_i, \mathbf{A}_i, s_i)$ with $i = \sigma^{-1}(j)$. In other words the discretized coordinates of Z' agree with the discretized coordinates of Z, up to permutation and possible axis reversal.

Main result: *This means that from observing X only, it suffices to find (learn) a g such that it yields a $p_{Z'}$ whose non-removable discontinuities forms some axis-aligned grid, for the resulting discretized coordinates of learned Z' to match the discretized coordinates of unobserved Z.*

4. Discussion and future work

In this work, we have proved that a relaxed form of *fully unsupervised identifiability* of latent factors is possible under *the most general nonlinear smooth mapping*, diffeomorphisms, a setup dominated by impossibility results in the literature. We are able to achieve the identification of discretized factor coordinates, provided that we assume axis-aligned discontinuities in the latent factor's distribution, which will form a grid. This proposed novel form of identifiability is meant as a step towards more realistic assumptions for disentanglement: no restrictive inductive bias on the mapping, no independence of factors, only potential causal footprints.

However, there are important limitations to this theory, the most obvious being that it requires actual *discontinuities*. This is required due to the flexibility of general diffeomorphisms. Future work shall try to relax this to just sharp (but not infinitely sharp) changes in the density, under slightly less general Lipschitz smooth mappings.

[1]Precisely, q is defined as $q(z_i, \mathbf{A}_i, 1) = \sum_{k=1}^{|\mathbf{A}_i|} \mathbf{1}_{z_i \geq \mathbf{A}_{i,k}}$ and $q(z_i, \mathbf{A}_i, -1) = |\mathbf{A}_i| - q(z_i, \mathbf{A}_i, 1)$.

When moving to the finite sample setting, we must resort to density estimation, which will yield a smoothed estimate of p_X, and as a result, discontinuities will become non-infinite sharp changes. These "softer" discontinuities can still be detected by considering the magnitude of the density gradient (as we display in Fig. 2). We discuss and explore in the appendix possible avenues for developing a concrete training criterion to recover the latent grid.

Lastly, the partitioning of the space through axis-aligned density landmarks is a general principle. It could lend itself to alternate coordinate systems, other than a grid, to locate cells. For example, a tree-like partitioning could support separators that do not necessarily split the entire space.

Acknowledgements

The authors thank Léon Bottou for sharing his original motivation of the appearance of density cliffs from a causal perspective, as well as David Lopez-Paz for related discussions on causal footprints. While, due to its identifiability focus, the present work only barely mentions these motivations, and certainly does not do them justice, the authors are very grateful for Léon's and David's encouragements in exploring this direction. The authors thank Diane Bouchacourt for discussions on Theorem 1 and on the experimental validation of the theory on real-world datasets, as well as Mark Ibrahim for his contribution to early discussions while this project was taking shape.

Vitória Barin-Pacela is partially supported by the Canada CIFAR AI Chair Program as well as a grant from Samsung Electronics Co., Ldt., administered by Mila, in support of her PhD studies at the University of Montreal.

Simon Lacoste-Julien and Pascal Vincent are CIFAR Associate Fellows in the Learning in Machines & Brains program.

This research has made use of the NASA Exoplanet Archive (Akeson et al., 2013), which is operated by the California Institute of Technology, under contract with the National Aeronautics and Space Administration under the Exoplanet Exploration Program.

References

Ahuja, K., Wang, Y., Mahajan, D., and Bengio, Y. Interventional causal representation learning. *arXiv preprint arXiv:2209.11924*, 2022.

Akeson, R. L., Chen, X., Ciardi, D., Crane, M., Good, J., Harbut, M., Jackson, E., Kane, S. R., Laity, A. C., Leifer, S., Lynn, M., McElroy, D. L., Papin, M., Plavchan, P., Ramí rez, S. V., Rey, R., von Braun, K., Wittman, M., Abajian, M., Ali, B., Beichman, C., Beekley, A., Berriman, G. B., Berukoff, S., Bryden, G., Chan, B., Groom, S., Lau, C., Payne, A. N., Regelson, M., Saucedo, M., Schmitz, M., Stauffer, J., Wyatt, P., and Zhang, A. The NASA exoplanet archive: Data and tools for exoplanet research. *Publications of the Astronomical Society of the Pacific*, 125(930):989–999, aug 2013. doi: 10.1086/672273. URL https://doi.org/10.1086%2F672273.

Bengio, Y., Courville, A., and Vincent, P. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798–1828, 2013.

Buchholz, S., Besserve, M., and Schölkopf, B. Function classes for identifiable nonlinear independent component analysis. *Conference on Neural Information Processing Systems*, 2022.

Comon, P. Independent component analysis, a new concept? *Signal processing*, 36(3):287–314, 1994.

Dittadi, A., Träuble, F., Locatello, F., Wuthrich, M., Agrawal, V., Winther, O., Bauer, S., and Schölkopf, B. On the transfer of disentangled representations in realistic settings. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=8VXvj1QNRl1.

do Carmo, M. P. *Differential Geometry of Curves and Surfaces*. Prentice-Hall, 1976.

Gresele, L., Von Kügelgen, J., Stimper, V., Schölkopf, B., and Besserve, M. Independent mechanism analysis, a new concept? *Advances in neural information processing systems*, 34:28233–28248, 2021.

Hyvärinen, A. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural Networks*, 10(3):626–634, 1999. doi: 10.1109/72.761722. URL https://doi.org/10.1109/72.761722.

Hyvärinen, A. and Morioka, H. Unsupervised Feature Extraction by Time-Contrastive Learning and Nonlinear ICA. In *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. URL https://proceedings.neurips.cc/paper/2016/file/d305281faf947ca7acade9ad5c8c818c-Paper.pdf.

Hyvärinen, A. and Morioka, H. Nonlinear ICA of Temporally Dependent Stationary Sources. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pp. 460–469, Fort Lauderdale, FL, USA, 20–22 Apr

2017. PMLR. URL http://proceedings.mlr.press/v54/hyvarinen17a.html.

Hyvärinen, A. and Pajunen, P. Nonlinear independent component analysis: Existence and uniqueness results. *Neural networks*, 12(3):429–439, 1999.

Hyvärinen, A., Sasaki, H., and Turner, R. E. Nonlinear ICA Using Auxiliary Variables and Generalized Contrastive Learning. In Chaudhuri, K. and Sugiyama, M. (eds.), *AISTATS*, volume 89 of *Proceedings of Machine Learning Research*, pp. 859–868. PMLR, 2019. URL http://dblp.uni-trier.de/db/conf/aistats/aistats2019.html#HyvarinenST19.

Khemakhem, I., Kingma, D., Monti, R., and Hyvarinen, A. Variational autoencoders and nonlinear ICA: A unifying framework. In *International Conference on Artificial Intelligence and Statistics*, pp. 2207–2217. PMLR, 2020.

Klain, D. A. and Rota, G.-C. *Introduction to Geometric Probability*. Cambridge University Press, 1997.

Lim, L.-H., Wong, K. S.-W., and Ye, K. The grassmannian of affine subspaces. *Foundations of Computational Mathematics*, 21:537—-574, 2021.

Locatello, F., Bauer, S., Lucic, M., Raetsch, G., Gelly, S., Schölkopf, B., and Bachem, O. Challenging common assumptions in the unsupervised learning of disentangled representations. In *international conference on machine learning*, pp. 4114–4124. PMLR, 2019.

Lopez-Paz, D., Nishihara, R., Chintalah, S., Schölkopf, B., and Bottou, L. Discovering causal signals in images. In *Proceedings of Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2017. URL http://leon.bottou.org/papers/lopezpaz-2017.

Roth, K., Ibrahim, M., Akata, Z., Vincent, P., and Bouchacourt, D. Disentanglement of correlated factors via hausdorff factorized support. *arXiv preprint arXiv:2210.07347*, 2022.

Taleb, A. and Jutten, C. Source separation in post-nonlinear mixtures. *IEEE Transactions on Signal Processing*, 47 (10):2807–2820, 1999. doi: 10.1109/78.790661.

Träuble, F., Creager, E., Kilbertus, N., Locatello, F., Dittadi, A., Goyal, A., Schölkopf, B., and Bauer, S. On disentangled representations learned from correlated data. In Meila, M. and Zhang, T. (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 10401–10412. PMLR, 18–24 Jul 2021.

Wang, Y. and Jordan, M. I. Desiderata for representation learning: A causal perspective. *arXiv preprint arXiv:2109.03795*, 2021.

APPENDIX

A. Practical criterion and Linear ICA comparison

A.1. Devising a practical criterion to recover an axis-aligned grid

We propose a practical criterion to learn the axis-aligned cliffs for illustrative purposes, as a proof of concept. Many other variants are conceivable. With finite samples, we use a density estimator $\hat{p}_{Z'}$, since we do not have access to the exact $p_{Z'}$. We remark that any density estimation will result in some smoothing of the true distribution. So even if there were real discontinuities in the exact density, they will appear as smoothed discontinuities: we will get density gradients with large magnitudes, not "infinite" magnitudes.

The steps for deriving a concrete training criterion are, thus, the following:

- Randomly initialize a parametric mapping $g : \mathcal{X} \to \mathcal{Z}$ to be learned.
- From observed (n, D), sample matrix $\mathbf{X}$, compute transformed (n, d) sample matrix $\mathbf{Z} = g(\mathbf{X})$ (g applied separately to each row). Note that we dropped the apostrophe ' in the $\mathbf{Z}$ to lighten notation for this practical criterion part.
- Build a density estimate, e.g. using a kernel-density-estimator (Parzen windows) $\hat{p}_\sigma$ and compute $\mathbf{V}_i = \frac{\partial \log \hat{p}_\sigma}{\partial z}(\mathbf{Z}_i)$ at every point of $\mathbf{Z}$
- Define an importance reweighting α based on the gradient magnitudes. e.g. $\alpha_i = \frac{\|V_i\|}{\sum_{i'} \|V_{i'}\|}$. This weight indicates, for each example, how close it is to a discontinuity or sharp density change (due to large density gradient magnitude). That is, how likely it belongs to an axis-separator of the grid.

From there, we can think of several terms that will encourage straightening and axis-aligning the points to shape them into an axis-aligned grid. As we aim to recover a smoothed version of the axis-aligned ground-truth grid, we try to make our weighted sample match different aspects of it. We can choose to align either the point samples, their gradient vectors, or both. Note that (smoothed) density gradient-vectors are also expected to be axis-aligned. Moreover, the alignment comes in two forms: local alignment in a neighborhood of points, and alignment to the axes. To encourage alignment of vectors, we maximize their cosine similarity $\mathrm{cosim}(\mathbf{a}, \mathbf{b}) = \frac{\langle \mathbf{a}, \mathbf{b} \rangle}{\|\mathbf{a}\|_2 \|\mathbf{b}\|_2}$. Let $\bar{\mathbf{V}}_i = \frac{\mathbf{V}_i}{\|\mathbf{V}_i\|}$ be the normalized version of $\mathbf{V}_i$. The following terms account for the desired alignment.

1. **Gradient local alignment term:** encourage pairs of neighboring points of high gradient magnitude to have gradients aligned by maximizing their cosine similarity. We can make the criterion a weighted average of cosine similarities (with significant weights only if they are neighboring points and have both large gradient magnitudes):

$$\begin{aligned}
\beta_{i,i'} =& \alpha_i \alpha_{i'} \exp(-\frac{1}{2\sigma_2^2} \|\mathbf{Z}_i - \mathbf{Z}_{i'}\|^2) \\
\bar{\beta}_{i,i'} =& \frac{\beta_{i,i'}}{\sum_{i,i'} \beta_{i,i'}} \\
\text{maximize} & \sum_{i,i'} \bar{\beta}_{i,i'} \,\mathrm{cosim}(\mathbf{V}_i, \mathbf{V}_{i'}) \\
i.e. \text{ minimize } \ell_{\mathrm{grad-local}} =& -\sum_{i,i'} \bar{\beta}_{i,i'} \left\langle \bar{\mathbf{V}}_i, \bar{\mathbf{V}}_{i'} \right\rangle
\end{aligned}$$

2. **Gradient axis alignment term:** encourages the individual gradient vectors to be axis-aligned, which can be obtained by maximizing the maximum cosine similarity with all the canonical axis vectors, which amounts to:

$$\text{maximize } \max_j |\mathrm{cosim}(V_i, \vec{\mathbf{1}}_j)| = \max_j \frac{|\mathbf{V}_{ij}|}{\|\mathbf{V}_i\|_2} = \frac{\|\mathbf{V}_i\|_\infty}{\|\mathbf{V}_i\|_2} = \|\bar{\mathbf{V}}_i\|_\infty$$

Over all the points, this would become

$$\text{minimize } \ell_{\text{grad-axis}} = -\sum_i \alpha_i \|\bar{\mathbf{V}}_i\|_\infty.$$

3. **Points local axis alignment term:** encourages neighboring points with large density gradient magnitude to lie on or close to the same axis separator. For this, it suffices that they share one of their coordinates. In other words, is suffices to minimize the minimum over coordinates of the squared difference:

$$\text{minimize } \ell_{\text{points-local}} = \sum_{i,i'} \bar{\beta}_{i,i'} \min_j \left(\frac{\mathbf{Z}_{ij} - \mathbf{Z}_{i'j}}{\|\mathbf{Z}_i - \mathbf{Z}_{i'}\|} \right)^2$$

4. **Points-gradient-orthogonality term:** encourages the gradient vector to be orthogonal to the vectors joining neighboring points by penalizing their squared cosine similarity, which adds the following term to the criterion:

$$\text{minimize } \ell_{\text{points-grad}} = \sum_{i,i'} \bar{\beta}_{i,i'} \left(\left\langle \bar{\mathbf{V}}_i, \frac{\mathbf{Z}_{i'} - \mathbf{Z}_i}{\|\mathbf{Z}_{i'} - \mathbf{Z}_i\|} \right\rangle \right)^2.$$

We can, then, define a training criterion that is a weighted sum of these terms (with appropriate sign), possibly together with the minimization of a reconstruction error ℓ_{rec} (from a decoder network $\hat{f}$ that tries to reconstruct $\mathbf{X}$ from $\mathbf{Z}$).

$$\ell_{\text{rec}} = \frac{1}{n} \sum_{i=1}^{n} \|\hat{f}(\mathbf{Z}_i) - \mathbf{X}_i\|^2$$

The complete loss to minimize is, thus,

$$L(\theta) = \lambda_1 \ell_{\text{grad-local}} + \lambda_2 \ell_{\text{grad-axis}} + \lambda_3 \ell_{\text{points-local}} + \lambda_4 \ell_{\text{points-grad}} + \lambda_5 \ell_{\text{rec}}$$

where θ is the set of (network) parameters of both encoder g and decoder $\hat{f}$.

A.2. Proof-of-concept experimentation

We study the case where the mixing model is linear for a comparison with Linear ICA. From the criterion proposed, we use only the gradient axis alignment term, since it is the most suitable for the linear case. We compare our model with Linear ICA and show that our model is able to learn a factorized representation of the factors, while Fast ICA (Hyvärinen, 1999) fails due to the correlation of the factors violating the independence assumption, as illustrated in Figure 3. The true factors have a correlation coefficient of 0.64.

Synthetic data generation: We generate a grid of points by establishing a prior for each cell, such that the sum of the priors of all the cells equals 1. We define a 4×4 grid, the position of each separator being drawn uniformly inside the range of the grid. In order to generate correlated data, first we draw the prior probabilities from a standard Uniform distribution. Then, we redefine the prior probability of the cells in the diagonal to be higher than the probability of the other cells, followed by normalization. The dataset is composed of 10,000 samples from this distribution.

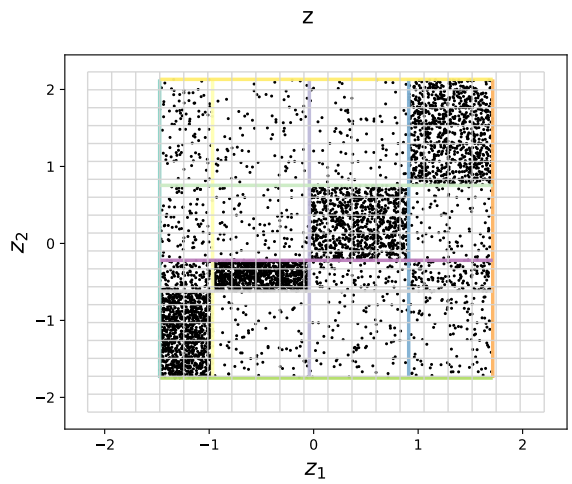
Experiment details: We minimize this loss using Stochastic Gradient Descent using a learning rate of 0.5 and a batch size of 5000 samples, which is half of the dataset. Mini-batches are employed due to the high memory cost of loading the full dataset. The optimization procedure runs for 200 epochs. The KDE estimation employs a bandwidth of 0.1.

Evaluation: The reconstruction of the factors is evaluated using the Mean Correlation Coefficient, which is standard in ICA literature.

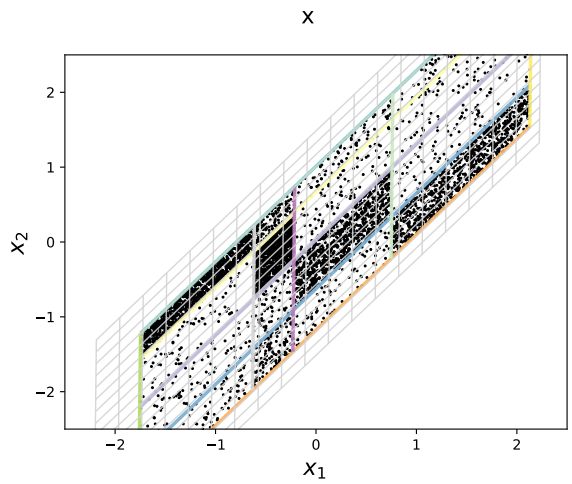
$$\text{MCC}(z, z') = \max_p \frac{1}{d} \sum_{i=1}^{d} \mid r(z_i, z'_{p[i]}) \mid, \tag{1}$$

where $z'_{p[i]}$ denotes a permutation of the variables of z', d is the number of factors, and $r(z_i, z'_i)$ computes the correlation coefficient of z_i and z'_i. It is maximized using the auction algorithm.

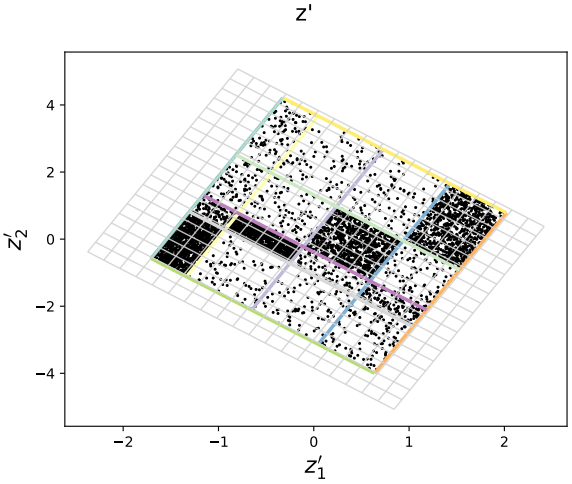
Our model obtains an MCC of 1.0, while FastICA obtains an ICA of 0.76. This is significant gain in performance reflects the factorization of the factors visualized in Figure 3.



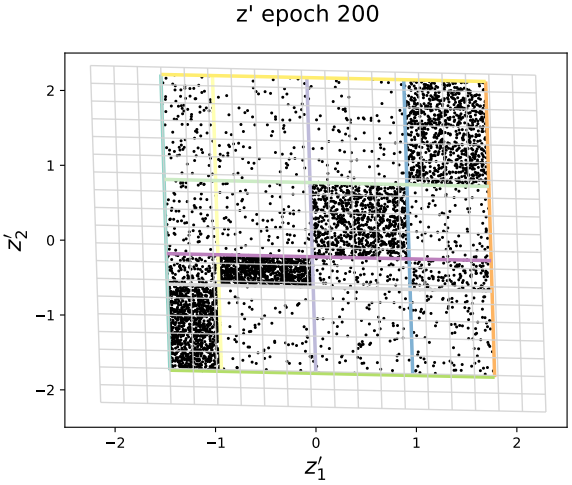
(a) True factors.



(b) Observed variables.



(c) Linear ICA reconstruction of the factors.



(d) Our model's reconstruction of the factors.

Figure 3: When the true latent factors (3a) are correlated, our method (3d) obtains a factorized representation corresponding to the ground-truth factors, as opposed to linear ICA (3c) which assumes independence of the factors and reconstructs the factors up to a rotation.

B. Main theorems

B.1. Non-removable discontinuity preservation theorem

The landmarks that we will show are preserved by diffeomorphism correspond to discontinuities in the PDF. One subtlety is that the PDF corresponding to a given distribution is not unique. The many PDFs representing a same distribution actually form an equivalence class, whose members may take arbitrarily different values on sets of points of measure zero. So not all discontinuities in a PDF are meaningful. Since we care about observable characteristics of the actual distribution, we must focus on aspects of the PDF that are immune to erasure by changes of measure zero. We use the following **definitions**:

Removable discontinuity: A PDF p has a removable discontinuity at z if p is discontinuous at z but there exists another p' in the same equivalence class (i.e. p and p' yield the exact same probability measure) that is continuous at z.

Non-removable discontinuity: A PDF p has a non removable discontinuity at z if p is discontinuous at z but that discontinuity is not removable. i.e. *all* PDFs in the equivalence class of p are discontinuous at z. Note that non-removable discontinuities are properties of an equivalence class of PDFs, thus of the distribution, not just of a single PDF.

Theorem 1. ***Non-removable discontinuity preservation theorem.*** *Let Z be a latent random variable with values in $\mathcal{Z} \subset \mathbb{R}^d$, whose distribution is represented by a PDF p_Z. Let $h : \mathcal{Z} \to \mathcal{Z}' \subset \mathbb{R}^d$ be a diffeomorphism, and let $Z' = h(Z)$ a transformed random variable whose distribution is represented by a probability density function $p_{Z'}$. Then $p_{Z'}$ has a non-removable discontinuity at a point z' if and only if p_Z has a non-removable discontinuity at point $z = h^{-1}(z')$.*

Proof. Let us denote $J_h(z) = \frac{\partial h}{\partial z}(z)$ the Jacobian of h, and $J_{h^{-1}}(z') = \frac{\partial h^{-1}}{\partial z'}(z')$ the Jacobian of h^{-1}. Suppose p_Z is one of the PDFs of Z, from this we can obtain a PDF of $Z^{'}$ using the change of variable formula: $p_{Z'}(z') = p_Z(h^{-1}(z'))|\det J_{h^{-1}}(z')|$. Symmetrically, we can say that if $p_{Z'}$ is a PDF of $Z^{'}$, we obtain a PDF version of Z as follows: $p_Z(z) = p_{Z'}(h(z))|\det J_h(z)|$.

Suppose the PDF p_Z has a non-removable discontinuity at z_0. Pick one of the PDFs of $Z^{'}$, let us call it $p_{Z'}$. There are three possibilities for what could happen at $h(z_0)$.

- $p_{Z'}$ is continuous at $h(z_0)$. We can apply the change of variables formula and obtain a PDF of Z that is given as $p^{'}_Z(z) = p_{Z'}(h(z))|\det J_h(z)|$. Since the RHS is made up of continuous functions, we conclude that $p^{'}_Z$ is continuous at z_0. This contradicts the fact that p_Z has a non-removable discontinuity at z_0.

- $p_{Z'}$ is discontinuous at $h(z_0)$ but the discontinuity is removable. Therefore, there exists a PDF $p^{'}_{Z'}$ that is continuous at $h(z_0)$. We can now follow the same argument as the above bullet to construct a PDF of Z that is continuous, which would contradict the fact that p_Z has a removable discontinuity at z_0.

- Finally, we are only left with the case that $p_{Z'}$ has a non-removable discontinuity at $h(z_0)$, which is what we set out to prove.

□

B.2. Grid structure recovery theorem and corollary

B.2.1. Definition of grid structure

The notions we use to define grid structure are related to usual hyperplanes and hypersurfaces of $\mathbb{R}^d$, but they are restricted to a connected subset $\mathcal{S}$ of $\mathbb{R}^d$ (which e.g. will be the support of the density on which density landmark grids can be defined, and may possibly not be defined outside).

Let $\mathcal{S} \subset \mathbb{R}^d$ be a connected smooth submanifold of dimension d ($\mathcal{S}$ could e.g. be an open $d-$ball), we will use the following **definitions**:

The **splitting** of a set $\mathcal{S}$ by another set $\mathcal{C}$, denoted $\text{split}(\mathcal{S}, \mathcal{C})$, is the set of *connected components* of $\mathcal{S} \setminus \mathcal{C}$. We say that $\mathcal{C}$ **splits** $\mathcal{S}$ **in two** to mean $|\text{split}(\mathcal{S}, \mathcal{C})| = 2$ (we denote the cardinality of a countable set A by $|A|$, and similarly for the number of elements in an ordered list or a tuple). We say that $\mathcal{C}$ is a **separator** of $\mathcal{S}$ if $\mathcal{C}$ is a *connected subset* of $\mathcal{S}$ and $\mathcal{C}$ splits $\mathcal{S}$ in two. The two connected components that result from the split are called the two **halves** resulting form the split,

denoted $\mathcal{C}^+$ and $\mathcal{C}^-$, i.e. $\{\mathcal{C}^+, \mathcal{C}^-\} = \text{split}(\mathcal{S}, \mathcal{C})$. $\mathcal{C}$ is a **smooth separator** of $\mathcal{S}$ if $\mathcal{C}$ is a *separator* of $\mathcal{S}$ and is a *smooth hypersurface* of $\mathcal{S}$ (i.e. a smooth embedded submanifold of dimension $d-1$).

An **axis-separator** of $\mathcal{S}$ is a special case of *smooth separator* of $\mathcal{S}$ that is the intersection of $\mathcal{S}$ with an *axis-aligned hyperplane* of $\mathbb{R}^d$ (a.k.a. a coordinate hyperplane). It can be defined as $\mathcal{H} = \Gamma_{\mathcal{S}}(i, \tau) = \{z \in \mathcal{S} | z_i = \tau\}$. Because it is a separator, it splits $\mathcal{S}$ in two halves $\Gamma^+_{\mathcal{S}}(i, \tau) = \{z \in \mathcal{S} | z_i > \tau\}$ and $\Gamma^-_{\mathcal{S}}(i, \tau) = \{z \in \mathcal{S} | z_i < \tau\}$, which are each nonempty and connected.

An **axis-separator-set** $\mathcal{G}$ on $\mathcal{S}$ is a set of axis separators.

An axis-aligned **grid** $G \subset \mathcal{S}$ is a subset of $\mathcal{S}$ that can be obtained as a union of all the separators in an axis-separator-set $\mathcal{G}$. i.e. $G = \cup\mathcal{G}$.
Note the important distinction we make between a *grid* which is a subset of $\mathcal{S}$, and hence a set of points, and an *axis-separator-set* which is a set of axis separators (which themselves are sets of points). An *axis-separator-set* thus has more explicit structure than a $grid$. The proof we will unroll depends conceptually on the ability to rebuild, in several steps, the entire grid internal structure, starting from only the unstructured *grid* as a set of points. The first step of this program will be the recoverability of *axis-separator-set* from *grid.*

A **parallel-separator-set** is a *set* of axis-separators all defined on the same i^{th} axis (and are thus parallel). In particular, we denote the subset of axis-separator set $\mathcal{G}$ that are all defined on the ith axis as $\mathcal{G}^{(i)}$.

A **discrete coordination** $\mathbf{A}$ is a tuple $\mathbf{A} = (\mathbf{A}_1, \ldots, \mathbf{A}_d)$ where each $\mathbf{A}_i$ is itself a tuple of real numbers in increasing order $\mathbf{A}_i = (\mathbf{A}_{i,1}, \ldots, \mathbf{A}_{i,n_i})$ s.t. $\mathbf{A}_{i,k+1} > \mathbf{A}_{i,k}$. These represent the coordinates of axis-separators along each of the d coordinate axes. A discrete coordination defines the entire grid structure. One can easily obtain the various constituent sets form it: a) individual separators ($\approx$"hyperplanes") are the $\Gamma_{\mathcal{S}}(i, \mathbf{A}_{i,k})$ and their positive and negative halves ($\approx$"half spaces"): $\Gamma^-_{\mathcal{S}}(i, \mathbf{A}_{i,k})$ and $\Gamma^+_{\mathcal{S}}(i, \mathbf{A}_{i,k})$; b) the *parallel-separator-sets* $\mathcal{G}^{(1)}, \ldots, \mathcal{G}^{(d)}$ are $\mathcal{G}^{(i)} = \text{parallelset}_{\mathcal{S}}(i, \mathbf{A}_i) = \{\Gamma_{\mathcal{S}}(i, \mathbf{A}_{i,k})\}_{k=1|}^{|\mathbf{A}_i|}$; c) the *axis-separator-set* $\mathcal{G} = \mathcal{G}^{(1)} \cup \ldots \cup \mathcal{G}^{(d)}$; and finally the grid $G = \cup\mathcal{G}$.

A **backbone** $\mathcal{H}^*$ of a grid is a *list* $\mathcal{H}^* = (\mathcal{H}^*_1, \ldots, \mathcal{H}^*_d)$ of d separators of that grid, each defined on the corresponding axis, that have a non-empty intersection (they meet at a single point z^*). $\mathcal{H}^*_1 \in \mathcal{G}^{(1)}, \ldots, \mathcal{H}^*_d \in \mathcal{G}^{(d)}$, $\bigcap_{i=1}^d \mathcal{H}^*_i = \{z^*\}$. For $\mathcal{H}^*$ to be a backbone, it is also required that $\mathcal{H}^*_i$ intersect *all* the other separators of the grid that are defined on the other axes (those not in the same parallel-separator-set): $\forall i, \forall j \neq i, \forall H \in \mathcal{G}^{(j)}, \mathcal{H}^*_i \cap H \neq \emptyset$. A backbone functions as a set of "main axes", and we require that a proper **grid** has at least one backbone. This is a weaker requirement than requiring a "complete grid" where *each* separator would be required to intersect all separators that are not in the same parallel-separator-set.

B.2.2. Grid structure preservation and recovery theorem

Theorem 2. ***Grid structure preservation and recovery theorem.*** *Suppose we have a smooth invertible mapping (diffeomorphism) $h : \mathbb{R}^d \to \mathbb{R}^d$, and an open connected subset $\mathcal{S} \subset \mathbb{R}^d$, we will denote its image through h as $\mathcal{S}' = h(\mathcal{S})$. Suppose we have an axis-aligned grid $G \subset \mathcal{S}$, associated with its axis-separator-set $\mathcal{G}$ and discrete coordination $\mathbf{A}$ i.e. $G = \text{grid}_{\mathcal{S}}(\mathbf{A})$. While the grid need not be "complete", we suppose $\mathcal{G}$ has at least one* backbone*. Now, suppose we have another axis-aligned grid in $\mathcal{S}'$, associated with its discrete coordination $\mathbf{B}$, i.e. $G' = \text{grid}_{\mathcal{S}'}(\mathbf{B})$. Suppose $G' = h(G)$. Then there exists a permutation function σ over dimension indexes $1, \ldots, d$ and a direction reversal vector $s \in \{-1, +1\}^d$ such that $\forall j \in \{1, \ldots, d\},\ i = \sigma^{-1}(j),\ K = |\mathbf{A}_i| = |\mathbf{B}_j|$, and $\forall k \in \{1, \ldots, K\}, \forall z' \in \mathcal{S}'$ we have*
If $s_i = +1$, then:

$$\begin{cases} z'_j = \mathbf{B}_{j,k} \iff h^{-1}(z')_i = \mathbf{A}_{i,k}, \\ z'_j > \mathbf{B}_{j,k} \iff h^{-1}(z')_i > \mathbf{A}_{i,k}, \\ z'_j < \mathbf{B}_{j,k} \iff h^{-1}(z')_i < \mathbf{A}_{i,k}; \end{cases}$$

If $s_i = -1$, then:

$$\begin{cases} z'_j = \mathbf{B}_{j,k} \iff h^{-1}(z')_i = \mathbf{A}_{i,K-k+1}, \\ z'_j > \mathbf{B}_{j,k} \iff h^{-1}(z')_i < \mathbf{A}_{i,K-k+1}, \\ z'_j < \mathbf{B}_{j,k} \iff h^{-1}(z')_i > \mathbf{A}_{i,K-k+1}. \end{cases}$$

Proof. See Section B.4 □

Corollary 3. ***Recovery of discretized coordinates.*** *We thus recover a (discretized version) of the ground truth coordinate system, with a one-to-one mapping of axes, and the ability from z' to distinguish whether the ground truth variable z was in one or the other specific cells of the ground-truth grid.*

More precisely we can define a quantization to integer of each coordinate z_i as:
$q(z_i, \mathbf{A}_i, 1) = \sum_{k=1}^{|\mathbf{A}_i|} \mathbf{1}_{z_i \geq \mathbf{A}_{i,k}}$ *and* $q(z_i, \mathbf{A}_i, -1) = |\mathbf{A}_i| - q(z_i, \mathbf{A}_i, 1)$ *in which case we have the coordinate equivalence:* $q(z'_j, \mathbf{B}_j, 1) = q(z_i, \mathbf{A}_i, s_i)$ with $i = \sigma^{-1}(j)$. *In other words* we have recovered quantized coordinates *up to permutation σ of the axes and possible direction reversal indicated by s.*

If we defined quantized coordinates $\bar{z}_i = q(z_i, \mathbf{A}_i, 1)$ and $\bar{z}'_i = q(z'_i, \mathbf{B}_i, 1)$ then $\bar{z}'_i = \bar{z}_{\sigma(i)}$ if $s_i = +1$ and $\bar{z}'_i = |\mathbf{B}'_i| - \bar{z}_{\sigma(i)}$ if $s_i = -1$.

B.3. Discretized coordinates identifiability theorem

Theorem 4. ***Discretized coordinates identifiability theorem.*** *Let Z be a latent random variable with values in $\mathcal{Z} \subset \mathbb{R}^d$ and whose PDF is p_Z. Let $f : \mathcal{Z} \to \mathcal{X} \subset \mathbb{R}^D$ be a diffeomorphism, and $X = f(Z)$ be the observed random variable. Further assume that the PDF p_Z has non-removable discontinuities that forms an axis-aligned grid, whose discrete coordination is $\mathbf{A}$. There exists diffeomorphisms $g : \mathcal{X} \to \mathcal{Z}'$ yielding a variable $Z' = g(X)$ whose PDF $p_{Z'}$ has non-removable discontinuities that form an axis-aligned grid. Consider any such diffeomorphism g, and let $\mathbf{B}$ the discrete coordination of its resulting axis-aligned grid. Then there exists a permutation function σ over dimension indexes $1, \ldots, d$ and a direction reversal vector $s \in \{-1, +1\}^d$ such that $q(Z'_j, \mathbf{B}_j, 1) = q(Z_i, \mathbf{A}_i, s_i)$ with $i = \sigma^{-1}(j)$. In other words the discretized coordinates of Z' agree with the discretized coordinates of Z, up to permutation and possible axis reversal.*

Main result: *This means that from observing X only, it suffices to find (learn) a g such that it yields a $p_{Z'}$ whose non-removable discontinuities forms some axis-aligned grid, for the resulting discretized coordinates of learned Z' to match the discretized coordinates of unobserved Z.*

Proof. Note that existence is trivial (it suffices to take $g = h^{-1}$ which yields $Z' = Z$). But the fact that *any* g that yields a PDF whose non-removable discontinuities form an axis-aligned grid will have this property can now easily be proven from our previous results. It suffices to consider $h = g \circ f$ which is a diffeomorphism (as the composition of two diffeomorphism) so that $Z' = h(Z)$ and to combine the non-removable discontinuity preservation theorem (Thm. 1) with the grid structure preservation and recovery theorem (Thm. 2). Let $G = \text{grid}_{\mathcal{S}}(\mathbf{A})$ and $G' = \text{grid}_{\mathcal{S}}(\mathbf{B})$ be the set of non-removable discontinuities of p_Z and $p_{Z'}$, respectively. From the non-removable discontinuity preservation theorem, we have that $G' = h(G)$. And from the grid structure preservation and recovery theorem and its corollary, we have that $G' = h(G)$ implies that there exists a permutation function σ over dimension indexes $1, \ldots, d$ and a direction reversal vector $s \in \{-1, +1\}^d$ such that $q(Z'_j, \mathbf{B}_j, 1) = q(Z_i, \mathbf{A}_i, s_i)$ with $i = \sigma^{-1}(j)$. We have thus proved that the discretized coordinates of Z' agree with the discretized coordinates of Z, up to permutation and axis reversal. □

B.4. Proof of the grid structure preservation and recovery Theorem 2

This is the more technical and long proof for which we postponed its dedicated section to here.

B.4.1. Principle of the proof

Starting from the premise $G' = h(G)$, we know that h maps every point of G to a point of G'. The proof recovers the entire underlying grid *structure* in 3 steps:

Step-1 recovers a one-to-one mapping of the individual separators: $\mathcal{G}' = h(\mathcal{G})$.

Step-2 recovers the partition into subsets of parallel separators (each subset associated to an axis): $\mathcal{G}'^{(j)} = h(\mathcal{G}^{(i)})$ (with permutation $j = \sigma(i)$).

Step-3 shows that the ordering of the separators in a parallel-separators-set is preserved (up to possible order reversal): $[h(\Gamma_{\mathcal{S}}(i, \mathbf{A}_{i,1})), \ldots, h(\Gamma_{\mathcal{S}}(i, \mathbf{A}_{i,K}))] = [\Gamma_{\mathcal{S}'}(j, \mathbf{B}_{j,1}), \ldots, \Gamma_{\mathcal{S}'}(j, \mathbf{B}_{j,K})]$ or in reversed order $[h(\Gamma_{\mathcal{S}}(i, \mathbf{A}_{i,1})), \ldots, h(\Gamma_{\mathcal{S}}(i, \mathbf{A}_{i,K}))] = [\Gamma_{\mathcal{S}'}(j, \mathbf{A}_{j,K}), \ldots, \Gamma_{\mathcal{S}'}(j, \mathbf{A}_{j,1})]$. And similarly for the halves corresponding to each of these separator. That a point belongs to a specific half allows us to tell whether the associated coordinate is above or below the associated threshold, which is what Theorem 2 expresses.

B.4.2. STEP 1 - RECOVERY OF ALL SEPARATORS

Knowing, from Theorem 1, that the set of *points* making up the axis-aligned grid G maps through h to the set of *points* making up the axis-aligned grid G' (i.e. $G' = h(G)$) our first major step consists in establishing that the *axis-separators* that make up G (i.e. the elements of $\mathcal{G}$) map one-to-one to the *axis-separators* that make up G' (i.e. the elements of $\mathcal{G}'$). We can denote this simply as $G' = h(G) \implies \mathcal{G}' = h(\mathcal{G})$.

PROOF FOR STEP 1

The high level proof is as follows (a complete detailed proof is provided in a further section) :

- Since $H \in \mathcal{G}$ is a connected smooth hypersurface in $\mathcal{S}$, and diffeomorphisms map connected sets to connected sets and smooth hypersurfaces to smooth hypersurfaces, we get that $h(H)$ is a connected smooth hypersurface in $\mathcal{S}'$.
- From Theorem 1, we also know that $h(H) \subset G'$.
- Next we establish that the only smooth connected hypersurfaces in $\mathcal{S}'$ that are included in G' are necessarily subsets of a single axis-separator of $\mathcal{G}'$. This is fundamentally due to the fact that a connected smooth hypersurface could not run along orthogonal intersections of the grid, as it would not be smooth (having a "kink"), so it has to stay within a single separator.
- We conclude that $h(H)$ is necessarily a subset of a single axis-separator $H' \in \mathcal{G}'$.
- We then show that not only is $h(H)$ a subset of a single axis-separator $H' \in \mathcal{G}'$, but that it has to be that entire separator. Because from the previous point, reverse diffeomorphism h^{-1} must map back the would-be remaining part of H' (i.e. $H' \setminus h(H) \neq \emptyset$) to a subset of the same separator as it maps back $h(H)$, i.e. to H. But this leads to a contradiction, since that remaining part did not come from H initially.
- We have thus shown that $H \in \mathcal{G} \implies h(H) \in \mathcal{G}'$. It suffices to apply this result in the other direction using h^{-1} to establish the converse. We thus have a bijection: the one-to-one mapping we needed to prove. Which we can write succinctly as $\mathcal{G}' = h(\mathcal{G})$.

B.4.3. STEP 2 - RECOVERY OF PARTITION INTO SETS OF PARALLEL SEPARATORS

We have established in step 1 that we recover the set of all separators $\mathcal{G}' = h(\mathcal{G})$. Our next step is to recover its partition into subsets of parallel separators (each subset associated to an axis): $\mathcal{G}'^{(j)} = h(\mathcal{G}^{(i)})$ (with permutation $j = \sigma(i)$).

PROOF FOR STEP 2

Consider d separators forming a backbone of $\mathcal{G}$, recall that a backbone is constituted of d distinct axis-separators that intersect in a single point, i.e. $\mathcal{H}_1^* \in \mathcal{G}^{(1)}, \dots, \mathcal{H}_d^* \in \mathcal{G}^{(d)}$, $\bigcap_{i=1}^d \mathcal{H}_i^* = \{z^*\}$.
We have that $\forall j \neq i,\ \mathcal{H}_i^* \neq \mathcal{H}_j^* \implies \forall j \neq i,\ h(\mathcal{H}_i^*) \neq h(\mathcal{H}_j^*)$.
We also have that $\bigcap_{i=1}^d \mathcal{H}_i^* = \{z^*\} \implies \bigcap_{i=1}^d h(\mathcal{H}_i^*) = \{h(z^*)\}$ (as h is a bijection).
Moreover, we know from step 1 that $\mathcal{H}_i^* \in \mathcal{G} \implies h(\mathcal{H}_i^*) \in \mathcal{G}'$. In short, the $h(\mathcal{H}_1^*), \dots, h(\mathcal{H}_d^*)$ are d distinct separators, each element of $\mathcal{G}'$, that intersect in a single point $h(z^*)$. The only sets of d *distinct* separators in $\mathcal{G}'$ that pass through a same point are d separators defined along each of the d different axes of $\mathcal{Z}' = \mathbb{R}^d$. Thus there exists a permutation σ such that for such backbone separators, $\mathcal{H}_i^* \in \mathcal{G}^{(i)} \implies h(\mathcal{H}_i^*) \in \mathcal{G}'^{(\sigma(i))}$.

Now consider any other separator $H \in \mathcal{G}^{(i)}$. From the definition of the backbone, we know that $H \cap \mathcal{H}_j^* \neq \emptyset, \forall j \neq i$.
This implies that $h(H) \cap h(\mathcal{H}_j^*) \neq \emptyset, \forall j \neq i$. The fact that $h(H)$ intersects a separator $h(\mathcal{H}_j^*) \in \mathcal{G}'^{(\sigma(j))}$ implies that it does not belong to parallel-separator-set $\mathcal{G}'^{(\sigma(j))}$. Thus $\forall j \neq i,\ h(H) \notin \mathcal{G}'^{(\sigma(j))}$. So there is just one parallel separator set left which $h(H)$ can belong to: $h(H) \in \mathcal{G}'^{(\sigma(i))}$.
In short, we have proved that $H \in \mathcal{G}^{(i)} \implies h(H) \in \mathcal{G}'^{(\sigma(i))}$.

Since distinct separators map to distinct separators, and each has to belong to exactly one of the $\mathcal{G}'^{(k)}$, this mapping is a bijection and we can write $H \in \mathcal{G}^{(i)} \iff h(H) \in \mathcal{G}'^{(\sigma(i))}$, or in short $h(\mathcal{G}^{(i)}) = \mathcal{G}'^{(j)}$ with $j = \sigma(i)$.

B.4.4. STEP 3 - RECOVERY OF COORDINATE ORDERING

The last step consists in showing that the ordering of the separators in a parallel-separators-set is preserved (up to possible order reversal).

PROOF FOR STEP 3

We only provide the high level principle here, the detailed proof can be found in the next section.

We first establish that h preserves separators and halves. This follows directly form the preservation of inclusion, connectedness and set operations under diffeomorphisms. Then we use the fact that inclusion defines a strict order relationship between positive halves associated to a coordination, and similarly between negative halves. As inclusion is preserved by a diffeomorphism, this order relationship is preserved. We can use this to show that the order implied by $\mathbf{A}_i$ is either conserved, as is, in $\mathbf{B}_j$ (negative halves of coordination $\mathbf{A}$ being mapped to negative halves of $\mathbf{B}$) or simply reversed (negative halves of $\mathbf{A}$ are being mapped to positive halves of $\mathbf{B}$). This directly yields the main Theorem 2 result.

C. Illustration of definitions

Figures 4, 5, and 6 illustrate the concepts used in the definitions of section B.2.

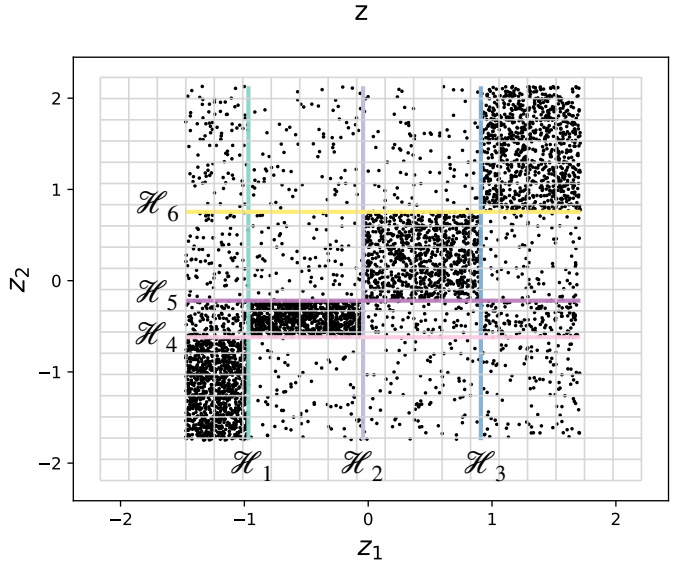


Figure 4: Axis separators $\mathcal{H}_1, \ldots, \mathcal{H}_6$ of $\mathcal{S}$.

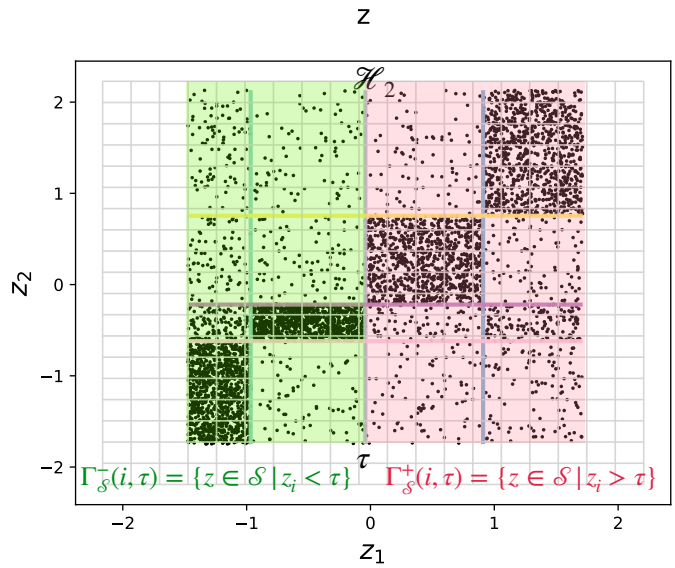


Figure 5: An **axis-separator** of $\mathcal{S}$ splits $\mathcal{S}$ in two halves $\Gamma^+_{\mathcal{S}}(i,\tau) = \{z \in \mathcal{S} | z_i > \tau\}$ and $\Gamma^-_{\mathcal{S}}(i,\tau) = \{z \in \mathcal{S} | z_i < \tau\}$, which are each nonempty and connected.

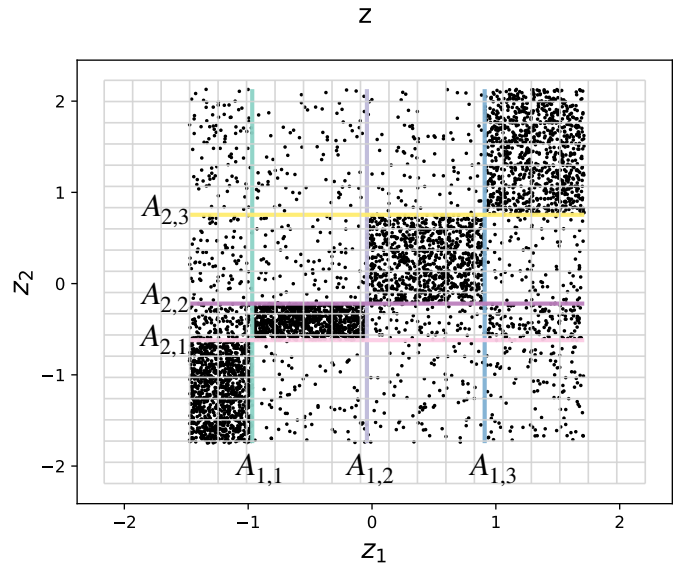


Figure 6: A **discrete coordination** $\mathbf{A}$ is a tuple $\mathbf{A} = (\mathbf{A}_1, \ldots, \mathbf{A}_d)$ where each $\mathbf{A}_i$ is itself a tuple of real numbers in increasing order $\mathbf{A}_i = (\mathbf{A}_{i,1}, \ldots, \mathbf{A}_{i,n_i})$ s.t. $\mathbf{A}_{i,k+1} > \mathbf{A}_{i,k}$. These represent the coordinates of axis-separators along each of the d coordinate axes.

D. Detailed Proofs

D.1. Detailed proof for Step 3

PRELIMINARY LEMMA

Lemma 1. *Preservation of separator and halves under diffeomorphism: If h is a diffeomorphism and $\mathcal{C}$ is a separator of $\mathcal{S}$ that splits it in two halves $\mathcal{C}^+$ and $\mathcal{C}^-$, then $h(\mathcal{C})$ is a separator of $h(\mathcal{S})$ that splits it in two halves $h(\mathcal{C}^+)$ and $h(\mathcal{C}^-)$*

Formally:

$$\begin{aligned}&\mathcal{C} \subset \mathcal{S}, \mathcal{C} \text{ connected}, \text{split}(\mathcal{S}, \mathcal{C}) = \{\mathcal{C}^+, \mathcal{C}^-\}\\ \Longleftrightarrow &h(\mathcal{C}) \subset h(\mathcal{S}), h(\mathcal{C}) \text{ connected}, \text{split}(h(\mathcal{S}), h(\mathcal{C})) = \{h(\mathcal{C}^+), h(\mathcal{C}^-)\}\end{aligned}$$

Proof. This follows from preservation of inclusion, connectedness, and set operations (union, intersection, difference) under a diffeomorphism.
Formally: $\mathcal{C} \subset \mathcal{S} \implies h(\mathcal{C}) \subset h(\mathcal{S})$.
$\mathcal{C}^+$ and $\mathcal{C}^-$ being the connected components of $\mathcal{S} - \mathcal{C}$ implies that $\mathcal{C}^+$ and $\mathcal{C}^-$ are each connected, and that $\mathcal{S} \setminus \mathcal{C} = \mathcal{C}^+ \cup \mathcal{C}^-$, where $\mathcal{C}^+ \cup \mathcal{C}^-$ is not connected.
Each of $\mathcal{S}$,$\mathcal{C}$, $\mathcal{C}^+$, $\mathcal{C}^-$ connected $\implies$ Each of $\mathcal{S}$, $h(\mathcal{C})$, $h(\mathcal{C}^+)$, $h(\mathcal{C}^-)$ connected.
$\mathcal{S} \setminus \mathcal{C} = \mathcal{C}^+ \cup \mathcal{C}^- \implies h(\mathcal{S}) \setminus h(\mathcal{C}) = h(\mathcal{C}^+) \cup h(\mathcal{C}^-)$
$\mathcal{C}^+ \cup \mathcal{C}^-$ not connected $\implies h(\mathcal{C}^+) \cup h(\mathcal{C}^-)$ not connected.
That $h(\mathcal{C}^+) \cup h(\mathcal{C}^-)$ is not connected but $h(\mathcal{C}^+)$ and $h(\mathcal{C}^-)$ are each connected, implies that $h(\mathcal{C}^+)$ and $h(\mathcal{C}^-)$ are the two connected components of $h(\mathcal{C}^+) \cup h(\mathcal{C}^-)$ i.e. of $h(\mathcal{S}) - h(\mathcal{C})$.
This implies that $\text{split}(h(\mathcal{S}), h(\mathcal{C})) = \{h(\mathcal{C}^+), h(\mathcal{C}^-)\}$. The implication in the other direction can be obtained in the by applying the same reasoning using h^{-1}. □

PROOF OF STEP 3

Let $j = \sigma(i)$ and $K = |\mathbf{A}_i| = |\mathbf{B}_j|$ and denote the corresponding set of axis separators as

$\mathcal{A} = \{\Gamma_{\mathcal{S}}(i, \mathbf{A}_{i,1}), \ldots, \Gamma_{\mathcal{S}}(i, \mathbf{A}_{i,K})\}$ and $\mathcal{B} = \{\Gamma_{\mathcal{S}'}(j, \mathbf{B}_{j,1}), \ldots, \Gamma_{\mathcal{S}'}(j, \mathbf{B}_{j,K})\}$

and denote the corresponding sets of halves:

$\mathcal{A}^+ = \{\Gamma^+_{\mathcal{S}}(i, \mathbf{A}_{i,1}), \ldots, \Gamma^+_{\mathcal{S}}(i, \mathbf{A}_{i,K})\}$, $\mathcal{A}^- = \{\Gamma^-_{\mathcal{S}}(i, \mathbf{A}_{i,1}), \ldots, \Gamma^-_{\mathcal{S}}(i, \mathbf{A}_{i,K})\}$, $\mathcal{A}^\pm = \mathcal{A}^+ \cup \mathcal{A}^-$

and $\mathcal{B}^+ = \{\Gamma^+_{\mathcal{S}'}(j, \mathbf{B}_{j,1}), \ldots, \Gamma^+_{\mathcal{S}'}(j, \mathbf{B}_{j,K})\}$, $\mathcal{B}^- = \{\Gamma^-_{\mathcal{S}'}(j, \mathbf{B}_{j,1}), \ldots, \Gamma^-_{\mathcal{S}'}(j, \mathbf{B}_{j,K})\}$, $\mathcal{B}^\pm = \mathcal{B}^+ \cup \mathcal{B}^-$

Proof Step 2, states that $h(\mathcal{A}) = \mathcal{B}$.

And we have from the above Lemma that

$$\begin{aligned}&\text{split}(\mathcal{S}, \mathcal{C}) = \{\mathcal{C}^+, \mathcal{C}^-\}\\ \Longleftrightarrow &\text{split}(h(\mathcal{S}), h(\mathcal{C})) = \{h(\mathcal{C}^+), h(\mathcal{C}^-)\}\end{aligned}$$

thus the equality of the sets of separators $h(\mathcal{A}) = \mathcal{B}$ obtained in Proof Step 2 implies an equality of the sets of halves:

$$h(\mathcal{A}^\pm) = \mathcal{B}^\pm$$

Now, the only halves, among all halves, that do not include any of the separators are $\Gamma^-_{\mathcal{S}}(i, \mathbf{A}_{i,1})$ and $\Gamma^+_{\mathcal{S}}(i, \mathbf{A}_{i,K})$ i.e. formally:

$$\{\mathcal{C} \in \mathcal{A} | \forall \mathcal{H} \in \mathcal{A}^\pm, \mathcal{C} \cap \mathcal{H} = \emptyset\} = \{\Gamma^-_{\mathcal{S}}(i, \mathbf{A}_{i,1}), \Gamma^+_{\mathcal{S}}(i, \mathbf{A}_{i,K})\}$$

this property will naturally translate to their mapping by diffeomorphism h (due to preservation of inclusion an intersections) hence

$$\{\mathcal{C} \in h(\mathcal{A}) | \forall \mathcal{H} \in h(\mathcal{A}^\pm), \mathcal{C} \cap \mathcal{H} = \emptyset\} = \{h(\Gamma^-_{\mathcal{S}}(i, \mathbf{A}_{i,1})), h(\Gamma^+_{\mathcal{S}}(i, \mathbf{A}_{i,K}))\}$$

i.e.

$$\{\mathcal{C} \in \mathcal{B} | \forall \mathcal{H} \in \mathcal{B}^\pm, \mathcal{C} \cap \mathcal{H} = \emptyset\} = \{h(\Gamma^-_{\mathcal{S}}(i, \mathbf{A}_{i,1})), h(\Gamma^+_{\mathcal{S}}(i, \mathbf{A}_{i,K}))\}$$

but we also have, similarly,

$$\{\mathcal{C} \in \mathcal{B} | \forall \mathcal{H} \in \mathcal{B}^{\pm}, \mathcal{C} \cap \mathcal{H} = \emptyset\} = \{\Gamma^{-}_{\mathcal{S}'}(i, \mathbf{B}_{i,1})), \Gamma^{+}_{\mathcal{S}'}(i, \mathbf{B}_{i,K}))\}$$

From this we conclude that:

$$\{h(\Gamma^{-}_{\mathcal{S}}(i, \mathbf{A}_{i,1})), h(\Gamma^{+}_{\mathcal{S}}(i, \mathbf{A}_{i,K}))\} = \{\Gamma^{-}_{\mathcal{S}'}(i, \mathbf{B}_{i,1})), \Gamma^{+}_{\mathcal{S}'}(i, \mathbf{B}_{i,K}))\}$$

Thus we have either one of two cases:

Case 1: $h(\Gamma^{-}_{\mathcal{S}}(i, \mathbf{A}_{i,1})) = \Gamma^{-}_{\mathcal{S}'}(i, \mathbf{B}_{i,1})$ and $h(\Gamma^{+}_{\mathcal{S}}(i, \mathbf{A}_{i,K})) = \Gamma^{+}_{\mathcal{S}'}(i, \mathbf{B}_{i,K})$. We associate this case with $s_i = +1$

Case 2: $h(\Gamma^{-}_{\mathcal{S}}(i, \mathbf{A}_{i,1})) = \Gamma^{+}_{\mathcal{S}'}(i, \mathbf{B}_{i,K})$ and $h(\Gamma^{+}_{\mathcal{S}}(i, \mathbf{A}_{i,K})) = \Gamma^{-}_{\mathcal{S}'}(i, \mathbf{B}_{i,1})$. We associate this case with $s_i = -1$

Case 1: $s_i = +1$, $h(\Gamma^{-}_{\mathcal{S}}(i, \mathbf{A}_{i,1})) = \Gamma^{-}_{\mathcal{S}'}(i, \mathbf{B}_{i,1})$ and $h(\Gamma^{+}_{\mathcal{S}}(i, \mathbf{A}_{i,K})) = \Gamma^{+}_{\mathcal{S}'}(i, \mathbf{B}_{i,K})$ The half-spaces in $\mathcal{A}^{\pm}$ that include $\Gamma^{-}_{\mathcal{S}}(i, \mathbf{A}_{i,1})$ are only the $\Gamma^{-}_{\mathcal{S}}$, formally:

$$\{\mathcal{H} \in \mathcal{A}^{\pm} | \Gamma^{-}_{\mathcal{S}}(i, \mathbf{A}_{i,1}) \subset \mathcal{H}\} = \mathcal{A}^{-}$$

this relationship will be maintained under diffeomorphism h i.e.

$$\{\mathcal{H} \in h(\mathcal{A}^{\pm}) | h(\Gamma^{-}_{\mathcal{S}}(i, \mathbf{A}_{i,1})) \subset \mathcal{H}\} = h(\mathcal{A}^{-})$$

thus, since $h(\mathcal{A}^{\pm}) = \mathcal{B}^{\pm}$ and $h(\Gamma^{-}_{\mathcal{S}}(i, \mathbf{A}_{i,1})) = \Gamma^{-}_{\mathcal{S}'}(i, B_{i,1})$ this can be rewritten as

$$\begin{aligned} \{\mathcal{H} \in \mathcal{B}^{\pm} | \Gamma^{-}_{\mathcal{S}'}(i, \mathbf{B}_{i,1}) \subset \mathcal{H}\} &= h(\mathcal{A}^{-}) \\ \mathcal{B}^{-} &= h(\mathcal{A}^{-}) \end{aligned}$$

or, written less compactly:

$$\{h(\Gamma^{-}_{\mathcal{S}}(i, \mathbf{A}_{i,1})), \ldots, h(\Gamma^{-}_{\mathcal{S}}(i, \mathbf{A}_{i,K}))\} = \{\Gamma^{-}_{\mathcal{S}'}(j, \mathbf{B}_{j,1}), \ldots, \Gamma^{-}_{\mathcal{S}'}(j, \mathbf{B}_{j,K})\}$$

Furthermore strict inclusion defines an order relationship between the elements of $\mathcal{A}^{-}$ which will be preserved under the diffeomorphism, and thus defines a strict ordering between them:

$$\begin{aligned} &\Gamma^{-}_{\mathcal{S}}(i, \mathbf{A}_{i,1}) \subsetneq \Gamma^{-}_{\mathcal{S}}(i, \mathbf{A}_{i,2}) \subsetneq \ldots \subsetneq \Gamma^{-}_{\mathcal{S}}(i, \mathbf{A}_{i,K}) \\ \implies &h(\Gamma^{-}_{\mathcal{S}}(i, \mathbf{A}_{i,1})) \subsetneq h(\Gamma^{-}_{\mathcal{S}}(i, \mathbf{A}_{i,2})) \subsetneq \ldots \subsetneq h(\Gamma^{-}_{\mathcal{S}}(i, \mathbf{A}_{i,K})) \end{aligned}$$

we know that the $h(\Gamma^{-}_{\mathcal{S}}(i, A_{i,k}))$ are the elements of $\mathcal{B}^{-}$ (as we have just sown that $\mathcal{B}^{-} = h(\mathcal{A}^{-})$), i.e. the $\Gamma^{-}_{\mathcal{S}'}(j, \mathbf{B}_{j,k})$. Their order is defined uniquely by strict inclusion as

$$\Gamma^{-}_{\mathcal{S}'}(j, \mathbf{B}_{j,1}) \subsetneq \Gamma^{-}_{\mathcal{S}'}(j, \mathbf{B}_{j,2}) \subsetneq \ldots \subsetneq \Gamma^{-}_{\mathcal{S}'}(j, \mathbf{B}_{j,K})$$

thus we can conclude not only (as we showed with $\mathcal{B}^{-} = h(\mathcal{A}^{-})$) that

$$\{h(\Gamma^{-}_{\mathcal{S}}(i, \mathbf{A}_{i,1})), \ldots, h(\Gamma^{-}_{\mathcal{S}}(i, \mathbf{A}_{i,K}))\} = \{\Gamma^{-}_{\mathcal{S}'}(j, \mathbf{B}_{j,1}), \ldots, \Gamma^{-}_{\mathcal{S}'}(j, \mathbf{B}_{j,K})\}$$

but also that their ordering is preserved i.e.

$$(h(\Gamma^-_{\mathcal{S}}(i, \mathbf{A}_{i,1})), \ldots, h(\Gamma^-_{\mathcal{S}}(i, \mathbf{A}_{i,K}))) = (\Gamma^-_{\mathcal{S}'}(j, \mathbf{B}_{j,1}), \ldots, \Gamma^-_{\mathcal{S}'}(j, \mathbf{B}_{j,K}))$$

or expressed differently:

$$\forall k \in \{1, \ldots, K\}, h(\Gamma^-_{\mathcal{S}}(i, \mathbf{A}_{i,k})) = \Gamma^-_{\mathcal{S}'}(j, \mathbf{B}_{j,k})$$

it is straightforward to conclude from this that we also have

$$\begin{aligned}
&\forall k \in \{1, \ldots, K\}, \\
&h(\Gamma^-_{\mathcal{S}}(i, \mathbf{A}_{i,k})) = \Gamma^-_{\mathcal{S}'}(j, \mathbf{B}_{j,k}) \\
&h(\Gamma^+_{\mathcal{S}}(i, \mathbf{A}_{i,k})) = \Gamma^+_{\mathcal{S}'}(j, \mathbf{B}_{j,k}) \\
&h(\Gamma_{\mathcal{S}}(i, \mathbf{A}_{i,k})) = \Gamma_{\mathcal{S}'}(j, \mathbf{B}_{j,k})
\end{aligned}$$

or stated differently, that:

$$\begin{aligned}
&\forall k \in \{1, \ldots, K\}, \forall z \in \mathcal{S} \\
&z \in \Gamma^-_{\mathcal{S}}(i, \mathbf{A}_{i,k}) \iff h(z) \in \Gamma^-_{\mathcal{S}'}(j, \mathbf{B}_{j,k}) \\
&z \in \Gamma^+_{\mathcal{S}}(i, \mathbf{A}_{i,k}) \iff h(z) \in \Gamma^+_{\mathcal{S}'}(j, \mathbf{B}_{j,k}) \\
&z \in \Gamma_{\mathcal{S}}(i, \mathbf{A}_{i,k}) \iff h(z) \in \Gamma_{\mathcal{S}'}(j, \mathbf{B}_{j,k})
\end{aligned}$$

or equivalently

$$\begin{aligned}
&\forall k \in \{1, \ldots, K\}, \forall z' \in \mathcal{S}', \\
&h^{-1}(z') \in \Gamma^-_{\mathcal{S}}(i, \mathbf{A}_{i,k}) \iff z' \in \Gamma^-_{\mathcal{S}'}(j, \mathbf{B}_{j,k}) \\
&h^{-1}(z') \in \Gamma^+_{\mathcal{S}}(i, \mathbf{A}_{i,k}) \iff z' \in \Gamma^+_{\mathcal{S}'}(j, \mathbf{B}_{j,k}) \\
&h^{-1}(z') \in \Gamma_{\mathcal{S}}(i, \mathbf{A}_{i,k}) \iff z' \in \Gamma_{\mathcal{S}'}(j, \mathbf{B}_{j,k})
\end{aligned}$$

which we may also write

$$\begin{aligned}
&\forall k \in \{1, \ldots, K\}, \forall z' \in \mathcal{S}', \\
&z'_j < \mathbf{B}_{j,k} \iff h^{-1}(z')_i < \mathbf{A}_{i,k} \\
&z'_j > \mathbf{B}_{j,k} \iff h^{-1}(z')_i > \mathbf{A}_{i,k} \\
&z'_j = \mathbf{B}_{j,k} \iff h^{-1}(z')_i = \mathbf{A}_{i,k}
\end{aligned}$$

which is what we needed to prove in the main grid structure recovery theorem.

Case 2: axis reversal $s_i = -1$, $h(\Gamma^-_{\mathcal{S}}(i, \mathbf{A}_{i,1})) = \Gamma^+_{\mathcal{S}'}(i, \mathbf{B}_{i,K})$ and $h(\Gamma^+_{\mathcal{S}}(i, \mathbf{A}_{i,K})) = \Gamma^-_{\mathcal{S}'}(i, \mathbf{B}_{i,1})$ We can follow the exact same reasoning steps as in case 1, starting from $h(\Gamma^-_{\mathcal{S}}(i, \mathbf{A}_{i,1})) = \Gamma^+_{\mathcal{S}'}(i, \mathbf{B}_{i,K})$:

- to first show that $h(\mathcal{A}^-) = \mathcal{B}^+$ i.e.

$$\{h(\Gamma^-_{\mathcal{S}}(i, \mathbf{A}_{i,1})), \ldots, h(\Gamma^-_{\mathcal{S}}(i, \mathbf{A}_{i,K}))\} = \{\Gamma^+_{\mathcal{S}'}(j, \mathbf{B}_{j,1}), \ldots, \Gamma^+_{\mathcal{S}'}(j, \mathbf{B}_{j,K})\}$$

- then use the preservation of the order relation defined by inclusion of halves to establish that

$$(h(\Gamma^-_{\mathcal{S}}(i, \mathbf{A}_{i,1})), \ldots, h(\Gamma^-_{\mathcal{S}}(i, \mathbf{A}_{i,K}))) = (\Gamma^+_{\mathcal{S}'}(j, \mathbf{B}_{j,K}), \ldots, \Gamma^+_{\mathcal{S}'}(j, \mathbf{B}_{j,1}))$$

- thus that

$$\begin{aligned}
&\forall k \in \{1, \dots, K\}, \\
&h(\Gamma^-_{\mathcal{S}}(i, \mathbf{A}_{i,k})) = \Gamma^+_{\mathcal{S}'}(j, \mathbf{B}_{j,K-k+1}) \\
&h(\Gamma^+_{\mathcal{S}}(i, \mathbf{A}_{i,k})) = \Gamma^-_{\mathcal{S}'}(j, \mathbf{B}_{j,K-k+1}) \\
&h(\Gamma_{\mathcal{S}}(i, \mathbf{A}_{i,k})) = \Gamma_{\mathcal{S}'}(j, \mathbf{B}_{j,K-k+1})
\end{aligned}$$

- conclude that

$$\begin{aligned}
&\forall k \in \{1, \dots, K\}, \forall z' \in \mathcal{S}', \\
&z'_j > \mathbf{B}_{j,k} \Longleftrightarrow h^{-1}(z')_i < \mathbf{A}_{i,K-k+1} \\
&z'_j < \mathbf{B}_{j,k} \Longleftrightarrow h^{-1}(z')_i > \mathbf{A}_{i,K-k+1} \\
&z'_j = \mathbf{B}_{j,k} \Longleftrightarrow h^{-1}(z')_i = \mathbf{A}_{i,K-k+1}
\end{aligned}$$

  which is what we needed to prove in the main grid structure recovery theorem.

D.2. Detailed proof of Step 1 – recovery of all separators

The goal of step 1 is to establish that the *axis-separators* that make up G (i.e. the elements of $\mathcal{G}$) map one-to-one to the *axis-separators* that make up G' (i.e. the elements of $\mathcal{G}'$). We can denote this simply as $G' = h(G) \Longrightarrow \mathcal{G}' = h(\mathcal{G})$.

A succinct overview of of the proof was given in Section B.4.2 in the main text. We provide a detailed proof here. Note that we always assume finite axis-separator sets.

PRELIMINARIES

Whenever we say hypersurface, it is always defined as a $d-1$ dimensional regular submanifold embedded in d-dimensional *ambient space* $\mathcal{S} \subset \mathbb{R}^d$, where $\mathcal{S}$ is a d-dimensional connected open submanifold of $\mathbb{R}^d$. In our application $\mathcal{S}$ will be the interior of the support of the density we consider.

- **Definition: Intersection set.** given a grid $G = \cup\mathcal{G} = \cup_{H \in \mathcal{G}} H$, we define its *intersection set* $I(\mathcal{G})$ as the set of points that belong to intersections of 2 or more distinct separators of $\mathcal{G}$. Formally: $I(\mathcal{G}) = \cup_{H \in \mathcal{G}, H' \in \mathcal{G}, H' \neq H}(H \cap H')$.
- **Definition: Exclusive point.** We say that a point z is exclusive to a separator H of a grid $G = \cup\mathcal{G}$ if it belongs to H but does not belong to any other separator of the grid (i.e. it does not belong to I). Similarly we will say that a set is exclusive to a separator if all its elements are exclusive points of that separator. The set of points of a separator H that are exclusive to it will be denoted $\breve{H} = H \setminus I$.
- **Definition: Tangent space** we view the tangent spaces to hypersurfaces embedded in an ambient space included in $\mathbb{R}^d$ literally as affine subspaces of $\mathbb{R}^d$, i.e. we use the traditional view[2] of tangent space (do Carmo, 1976), which is a natural generalization of the notion of a plane tangent to a surface at a point, to higher dimensional hypersurfaces embedded in $\mathbb{R}^d$. The tangent space at $z \in A$ to a hypersurface A will be denoted $T^A(z) = T_z A$. A smooth hypersurface has the property that it has at every $z \in A$ a well-defined tangent space $T^A(z) = T_z A$ of the same dimension as the hypersurface. When A is a smooth hypersurface, T^A is a smooth map $T^A : A \to \mathrm{Graff}_{d-1}(\mathbb{R}^d)$ that maps any point z of A to a point of the *affine-Grassmannian manifold* (Klain & Rota, 1997; Lim et al., 2021) $\mathrm{Graff}_{d-1}(\mathbb{R}^d)$, i.e. the space of all $d-1$ dimensional affine-subspaces of $\mathbb{R}^d$. Since T^A is a continuous map between smooth manifolds, $T^A(z)$ will be continuous in any local (or global) parametrization of A around z. (continuity based on the topology of the affine-Grassmannian manifold for comparing tangent spaces as affine-subspaces of $\mathbb{R}^d$).
  Note that the tangent space to any axis-separator H is a constant: it is the subspace confounded with the hyperplane that includes the separator, and will be denoted $\mathcal{T}_H$. i.e. we have $\forall z \in H,\ T^H(z) = T_z H = \mathcal{T}_H$. Note also that with this affine subspace definition of tangent space, $\mathcal{T}_H$ is different for every separator H of an axis-aligned grid: $\forall H_1 \in \mathcal{G}, \forall H_2 \in \mathcal{G}, \mathcal{T}_{H_1} = \mathcal{T}_{H_2} \Leftrightarrow H_1 = H_2$.

[2] This traditional extrinsic view of tangent space is preferred here to more modern definitions, because it simplifies a step in our proof. It is also arguably easier to intuit and follow for readers who may not be familiar with differential geometry.

- **Useful properties:** We will also use the following properties that are either well-established differential geometry knowledge or straightforward corollaries thereof
  - **Property 1:** *A diffeomorphism maps a smooth hypersurface to a smooth hypersurface*
  - **Property 2:** *A diffeomorphism maps a path-connected set to a path-connected set.*
  - **Property 3:** *Smooth connected hypersurfaces in $\mathbb{R}^d$ have a $d-1$ dimensional tangent space that is well-defined all over the hypersurface and continuous (in the sense defined above, see Tangent space)*
  - **Property 4:** *A non-empty open subset of a smooth hypersurface in ambient space is itself a smooth hypersurface in ambient space*
  - **Property 5:** *A hypersurface that is a subset of another hypersurface has at every of its points the same tangent space as the hypersurface it is a subset of.*

DETAILED PROOF OF STEP 1

Lemma 2. *No subset of the intersection set $I(\mathcal{G})$ of an axis-aligned grid $G = \cup\mathcal{G}$ can be a hypersurface in ambient space.*

Proof. Consider $I(\mathcal{G})$ the intersection set of a grid $G = \cup\mathcal{G}$. Formally: $I(\mathcal{G}) = \cup_{H\in\mathcal{G},H'\in\mathcal{G},H'\neq H}(H \cap H')$. Each $H \cap H'$, if it is non-empty, is the intersection of two orthogonal (thus transversal) connected hypersurfaces (i.e. $d-1$ dimensional submanifolds embedded in ambient space), so that their intersection can be at most a $d-2$ dimensional embedded submanifold of ambient space. The union of a finite number of at most $d-2$ dimensional submanifolds cannot be more than $d-2$ dimensional, so $I(\mathcal{G})$ cannot be more than $d-2$ dimensional. Consequently no subset of $I(\mathcal{G})$ can be more than $d-2$ dimensional, thus it cannot be a hypersurface in ambient space. □

Lemma 3. *Let A be a connected smooth hypersurface included in an axis-alined grid $G = \cup\mathcal{G}$ with axis-separator set $\mathcal{G}$. Let $z \in A$. All open neighborhoods of z in A will necessarily contain at least one point that is exclusive to a separator of $\mathcal{G}$.*

Proof. An open neighborhood $\mathcal{B}_z^A$ of z in A is an open subset of A, thus from Property 4, $\mathcal{B}_z^A$ is a hypersurface in ambient space. From Lemma 2 no subset of $I(\mathcal{G})$, (the set of points of G that belong to more than one separator) can be a hypersurface. So $\mathcal{B}_z^A$ cannot be a subset of $I(\mathcal{G})$, i.e. it must contain at least one point exclusive to a separator of $\mathcal{G}$. □

Lemma 4. *Let A be a connected smooth hypersurface included in an axis-alined grid $G = \cup\mathcal{G}$ with axis-separator set $\mathcal{G}$. Let z be a point of A that is exclusive to a separator $H \in \mathcal{G}$ (i.e. $z \in H \setminus I(\mathcal{G})$: it belongs to no other separator of $\mathcal{G}$), then there exists an open connected neighborhood $\mathcal{B}_z^A$ of z in A that is exclusive to H.*

Proof. We reason using the usual Euclidean distance in $\mathbb{R}^d$. Consider an open d-ball $\mathcal{B}_z^d$ in $\mathbb{R}^d$ centered on z and whose radius ϵ is chosen to be less than the smallest distance of z to any other separator, i.e. such that $0 < \epsilon < \inf_{z'\in(G\setminus H)} \|z - z'\|$. Since z is exclusive to separator H and the number of separators is finite, this distance will be greater than 0. Then all points of G within a distance less than ϵ of z will necessarily belong exclusively to H, i.e. $\mathcal{B}_z^d \cap G \subset \breve{H}$, where $\breve{H} = H \setminus I(\mathcal{G})$. Now we can choose a sufficiently small connected open neighborhood $\mathcal{B}_z^A$ of z in A so that the distance in ambient space between z and any other point of $\mathcal{B}_z^A$ is less than ϵ. Thus $\mathcal{B}_z^A \subset \mathcal{B}_z^d$. Since we also have $\mathcal{B}_z^A \subset A \subset G$ this implies that $\mathcal{B}_z^A \subset \mathcal{B}_z^d \cap G$ and consequently that $\mathcal{B}_z^A \subset \breve{H}$. We have thus shown that there exists an open connected neighborhood of z in A that is exclusive to H. □

Lemma 5. *Let A be a connected smooth hypersurface included in an axis-alined grid $G = \cup\mathcal{G}$ with axis-separator set $\mathcal{G}$. Then for any point $z \in A$ there exists a non-empty open subset B whose boundary contains z and such that B is a non-empty open subset exclusive to one of the separators.*

Proof. There are two cases to consider for z: either z is an exclusive point of a separator of the grid, or it is an intersection point of separators (belonging to $I(\mathcal{G})$).
First case: z is a point exclusive to a separator $H \subset \mathcal{G}$.
Then by Lemma 4, we know that there exists an open connected neighborhood $\mathcal{B}_z^A$ of z in A that is exclusive to H. We can then easily pick an open subset B of $\mathcal{B}_z^A$ whose boundary contains z (For instance, pick a close neighbor z_1 of z in $\mathcal{B}_z^A$, and construct B as the intersection of $\mathcal{B}_z^A$ with an open ball centered on z_1 and of radius $\|z_1 - z\|$). B is an open subset exclusive to H, the separator that z belongs to.

Second case: z is not exclusive to any separator of the grid.
Let $\mathcal{G} = \{H_1, \dots, H_k\}$ the finite set of separators of grid $G = \cup\mathcal{G}$. Let $\breve{H}_i = H_i \setminus I(\mathcal{G})$ the corresponding subset of exclusive points to each separator H_i, and let $\breve{A}_i = \breve{H}_i \cap A$, for each $i \in \{1, \dots, k\}$. So $\breve{A}_i$, if it is not empty, will contain only points exclusive to H_i. From Lemma 4 we deduce that every point of $\breve{A}_i$ has an open neighborhood in A exclusive to H_i: this open neighborhood is thus included in $A \cap \breve{H}_i$ and is thus a subset of $\breve{A}_i$. We have thus shown that every point of $\breve{A}_i$ has an open neighborhood in A that is included in $\breve{A}_i$. From this we conclude that each $\breve{A}_i$ is an open subset (possibly empty) of A.
We know that z belongs to none of the $\breve{A}_i$, since it is not exclusive to any separator. Now we will show that z belongs to the *boundary* of at least one of the $\breve{A}_i$. We will reason using the metric d^A induced on embedded submanifold $A \subset \mathbb{R}^d$ by the usual Euclidean metric in ambient space $\mathbb{R}^d$. Let $\epsilon = \min_{i \in \{1,\dots,k\}} d^A(z, \breve{A}_i)$. We can use a min since it is over a finite number k of separators. Note that $d^A(z, \breve{A}_i) = \inf_{z' \in \breve{A}_i} d^A(z, z')$ will be $+\infty$ if $\breve{A}_i$ is empty, by the definition of the infimum. If ϵ was strictly greater than 0, then this would mean that no point of A exclusive to any separator would be at a distance strictly less than ϵ from z (since any point of A exclusive to a separator belongs to one of the $\breve{A}_i$). Thus the open ball $\mathcal{B}^A(z, \epsilon) = \{z' \in A, d^A(z, z') < \epsilon\}$ would not contain any point exclusive to any separator. But this would contradict Lemma 3. So necessarily $\epsilon = 0$. This implies that there is at least one of the $\breve{A}_i$ whose distance to z is 0, i.e. there exists a $k^* \in \{0, \dots, k\}$ such that $d^A(z, \breve{A}_{k^*}) = 0$. Since $z \notin \breve{A}_{k^*}$ we conclude that z belongs to the *boundary* of this $\breve{A}_{k^*}$. Moreover this $\breve{A}_{k^*}$ is non-empty (otherwise that distance would be $+\infty$). It is thus an open-subset of A, exclusive to separator H_k. We have thus established that there exists a non-empty open subset of A exclusive to one of the separators, and whose boundary contains z.

□

Lemma 6. *Let A be a connected smooth hypersurface included in an axis-alined grid $G = \cup\mathcal{G}$ with axis-separator set $\mathcal{G}$. Let γ be a continuous path, included in A, that starts at a point z_1, where z_1 is exclusive to a separator $H \in \mathcal{G}$. Then γ will necessarily be included entirely in H.*

Proof. Consider path $\gamma : [0, 1] \to A$, where $\gamma(0) = z_1$ is exclusive to H. From Lemma 5, for each point $\gamma(t) \in A$, there exists an open subset B_t of A whose boundary contains $\gamma(t)$ and such that B_t is an open subset exclusive to one of the separators. Let us call this separator H_t (Note that there may be multiple possible choices for B_t and H_t). Consider any point $z \in B_t$. Since B_t is a non-empty open subset of smooth hypersurface A, by Property 4 it is a hypersurface and by Property 5 B_t and A will have the same tangent space, so that $T_zB_t = T_zA$. Since B_t is also a subset of smooth hypersurface H_t, we have by Property 5 that $T_zB_t = T_zH_t$. Thus $T_zA = T_zB_t = T_zH_t$. Now the tangent space to any axis-separator H' is the constant $\mathcal{T}_{H'}$. We can thus write, for any $z \in B_t$, $T_zA = T_zB_t = T_zH_t = \mathcal{T}_{H_t}$. Since A is a connected smooth hypersurface, it has a continuous and well defined tangent space at every point. Thus if the tangent space is constant on an open subset $B_t \subset A$ it will have that same constant value at its boundary. So the tangent space to A at point $\gamma(t)$, which belongs to the boundary of B_t, will also be $T_{\gamma(t)}A = \mathcal{T}_{H_t}$. For the same reason of the continuity of the tangent space of a path-connected smooth hypersurface A, we cannot have, along the curve $\gamma(t)$, an abrupt change in the tangent space $T_{\gamma(t)}A$, consequently $\mathcal{T}_{H_t}$ cannot change abruptly along the path. The only way for it not to change abruptly is that H_t stays constant along the path: $H_t = \text{constant}\ \forall t \in [0, 1]$. In other words, for any point $\gamma(t)$ along the path there must exist an open subset B_t of A that is included in and exclusive to the same constant separator along the path. Now, if $z_1 = \gamma(0)$ is exclusive to a separator H, then $H_0 = H$ and we must thus have $H_t = H, \forall t \in [0, 1]$. We have thus shown that if the path starts at a point $z_1 = \gamma(0)$ which is exclusive to a separator H, then all points of the path necessarily belong to H (though not necessarily exclusively to H). Thus, the path is entirely included in H. □

Lemma 7. *A path-connected smooth hypersurface A included in an axis-alined grid $G = \cup\mathcal{G}$ is necessarily a subset of one separator of $\mathcal{G}$.*

Proof. Let z_0 be a point of A that is exclusive to a separator $H \in \mathcal{G}$. We know from Lemma 3 that such a point exists. Since A is path-connected, there exists in A a continuous path connecting z_0 to any point $z \in A$. Thus, Lemma 6 leads to conclude that $\forall z \in A, z \in H$. Thus $A \subset H$. □

Lemma 8. Let $G = \cup\mathcal{G}$ be an axis-aligned grid in $\mathcal{S}$. Let $h : \mathcal{S} \to \mathcal{S}'$ be a diffeomorphism. Let $G' = \cup\mathcal{G}'$ be an axis-aligned grid in $\mathcal{S}'$. If $h(G) = G'$, the image of a separator $H_1 \in \mathcal{G}$ by the diffeomorphism h will be a separator $H' \in \mathcal{G}'$, i.e. $H_1 \in \mathcal{G} \implies h(H_1) \in \mathcal{G}'$.

Proof. An axis-separator $H_1 \subset G$ is a path-connected smooth hypersurface. From Property 1 and Property 2, its image by diffeomorphism h will be a path-connected smooth hypersurface $h(H_1) \subset G'$. So $h(H_1)$ is a path-connected smooth hypersurface included in axis-aligned grid G'. Consequently, Lemma 7 guarantees that we have $h(H_1) \subset H'$ for some $H' \in \mathcal{G}'$. We will now prove that $h(H_1) = H'$. Suppose by contradiction that $h(H_1) \subsetneq H'$, and let $B' = H' - h(H_1) \neq \emptyset$. Similarly, if we apply the reverse diffeomorphism h^{-1}, we will have $h^{-1}(H') \subset H_2$ for some $H_2 \in \mathcal{G}$. Consequently, the two disjoint sets composing $H' = B' \cup h(H_1)$ will both map back to subsets of H_2, i.e. $h^{-1}(B') \subset H_2$ and $h^{-1}(h(H_1)) \subset H_2$. The latter can be rewritten as $H_1 \subset H_2$, which implies $H_2 = H_1$ since no two distinct separators of $\mathcal{G}$ are included in one another. So we have $h^{-1}(B') \subset H_1$. Thus $h(h^{-1}(B')) \subset h(H_1)$, hence $B' \subset h(H_1)$. We had defined B' as $B' = H' - h(H_1) \neq \emptyset$ but a non-empty B' cannot at the same time correspond to a set from which we removed $h(H_1)$ *and* be included in $h(H_1)$. We have a contradiction, so we cannot have $h(H_1) \subsetneq H'$, therefore $h(H_1) = H'$. □

Proposition 9. *The diffeomorphism h maps separators in $\mathcal{G}$ one-to-one to separators in $\mathcal{G}'$, i.e. $h(G) = G' \implies h(\mathcal{G}) = \mathcal{G}'$.*

Proof. We have shown in Lemma 8 that $H \in \mathcal{G} \implies h(H) \in \mathcal{G}'$. It suffices to apply this result in the other direction using h^{-1} to establish the converse. We thus have a bijection: the one-to-one mapping we needed to prove. Which we can write succinctly $h(\mathcal{G}) = \mathcal{G}'$. □