International Institute for
Applied Systems Analysis
IIASA www.iiasa.ac.at

On Multivariate Structures and Exhaustive Reductions

Chiaromonte, F.

IIASA Interim Report
November 1998

IIASA

International Institute for Applied Systems Analysis • A-2361 Laxenburg • Austria
Tel: +43 2236 807 • Fax: +43 2236 71313 • E-mail: info@iiasa.ac.at • Web: www.iiasa.ac.at

INTERIM REPORT IR-98-080 / September 1998

On Multivariate Structures and Exhaustive Reductions

Francesca Chiaromonte (chiaro@iiasa.ac.at)

Approved by
Giovanni Dosi (dosi@iiasa.ac.at)
Leader, _TED Project_

Abstract

Simplified representations of multivariate laws, and in particular those allowing one to decrease the dimension while preserving structural information, are of paramount importance in statistical analysis. This paper concerns the *theoretical premises* of simplification. We introduce a framework that allows us to specify *definitions of structure* for multivariate laws. Conceiving definitions as *partitions* of the probability laws on a Euclidean space, we show how they can be generated via *partial orders*, or *binary operations* and *noise classes.* Moreover, the framework allows us to identify *simplified representations* that are guaranteed to be *exhaustive* with respect to such definitions, and might live in *lower dimension.*

Acknowledgements

I would like to thank John R. Baxter and R. Dennis Cook for the valuable help they have provided throughout the development of this work. I would also like to thank Pietro Muliere and Piercesare Secchi for their comments and insights.

Contents

On Multivariate Structures and Exhaustive Reductions

Francesca Chiaromonte (chiaro@iiasa.ac.at)

Introduction

A k-variate law is a complex object whose structure embodies both marginal and joint features. Dimension does not affect the analysis of marginal features, but as k increases it becomes progressively harder to conceive and articulate the joint ones [1]. Furthermore, any dimension exceeding 3 prevents a direct visualization of data drawn from the law, impairing the graphical preliminary exploration that ought to precede any model-based analysis.

As very high-dimensional data sets become more frequent in several fields of application, the above considerations motivate the quest for *simplified representations* of multivariate laws, especially ones that induce a *drop in dimension.* Since these representations are aimed at capturing structural traits, their degree of informativeness with respect to the structure becomes an issue. In particular, one might want to develop simplified representations that are *exhaustive*, i.e. that entirely preserve structural information.

At the population level, many of the graphical exploration procedures developed in the last decades can be interpreted as low-dimensional projective representations of the law from which the data are drawn. They can be simple –*views*– or multiple –*tours*, and their projection subspaces are usually selected targeting some quantitative proxy for the structure. To mention some examples, *Principal Components Analysis* targets variability, while several versions of *Projection Pursuit* target indexes expressing departure from normality ("non-linear" structure of the law). *Grand Tours* target geometric features that, loosely speaking, are retained in low dimension with high probability (consequently, the subspaces are selected at random). Some recent and very sophisticated computer tools combine those proxies and allow the user to perform *multiple tours* which are partly random, partly guided by indexes, and partly guided by the user himself (see A. Buja, D. Cook, D.F. Swayne, 1996 –and references therein).

Articulating and multiplying proxies, as well as augmenting the size (taking longer tours), clearly increase the information content of the representation with respect to the structure. On the other hand, informativeness and exhaustiveness of a view or tour remain imponderable unless one grounds such representations theoretically by means of a suitable *definition of structure*, and a *reduction scheme* to go with it. An example of this approach is *Factor Analysis*: Structure is defined as linear interdependence, i.e. pair-wise correlations among coordinate components. The reduction scheme is given by an additive decomposition model separating a latent factor, which exhaustively embodies correlations,

[1] Regarding interdependence among coordinate components, one key to its articulation is *conditional independence* (see A.P. Dawid, 1979). A very interesting representation of interdependencies through *conditional independence graphs* is given by J. Whittaker, 1990.

from a noise that does not contribute to them. Thus, the noise can be neglected and the analysis restricted to the smallest subspace supporting the latent factor [2].

Our focus in this paper is not on techniques to make inference on simplified representations (population objects) based on data from a multivariate law, but on the theoretical premises for simplification. We introduce a framework that allows us to specify definitions of multivariate structure, and reduction schemes that identify exhaustive simplified representations. Such a framework does not rely on strong requirements on the nature of the law under investigation.

Let $\mathcal{L}$ be the collection of all probability laws on $\mathbb{R}^k$. One can think of a definition of structure for k-variate laws as a *partition* $\mathcal{C}$ of $\mathcal{L}$; laws in a class C are meant to share a common structure (i.e. to be structurally equivalent). Conceiving definitions of structure as partitions has an immediate advantage: Once a definition/partition is chosen, any member of the class C to which a law L belongs can be taken in lieu of L itself in order to investigate its structure. In other words, *laws in C provide alternative representations of L which exhaustively preserve structural information.* Some such laws might allow to observe the structure *more easily or clearly.* In particular, some such laws might *live in algebraic dimension smaller than k*, even though the original L does not.

In Section 1, we show how a definition/partition can be generated via an appropriate *partial order.* We also introduce and characterize the basic components of our framework: *origins*, the *reduction operator*, *invariances*, *invariant marginalizations* and *kernels.* In Section 2, we see how a definition/partition can be generated via an appropriate *binary operation* and *noise class*, and further characterize the basic components. Section 3 contains an example of a definition based on a binary operation and a noise class; namely *convolution and white noises.* Some final remarks are given in Section 4.

1 Generating partitions via partial orders

A partition of $\mathcal{L}$ can be generated via an appropriate partial order. Let $\preceq$ be a partial order on $\mathcal{L}$; that is, a relation such that

- $P(1)$: $\forall L \in \mathcal{L}$, $L \preceq L$ (reflexivity)
- $P(2)$: if $L_1 \preceq L_2$ and $L_2 \preceq L_3$, then $L_1 \preceq L_3$ (transitivity)
- $P(3)$ if $L_1 \preceq L_2$ and $L_2 \preceq L_1$, then $L_1 = L_2$ (antisymmetry).

A **chain** of $\preceq$ is defined as a subset of $\mathcal{L}$ restricted to which $\preceq$ becomes a complete order; that is, a subset containing only comparable elements. A **maximal chain** C of $\preceq$ is defined as a chain which is not a proper subset of any other chain; that is, a subset obtained by grouping together *all* elements comparable to each other. We indicate the collection of maximal chains with $\mathcal{C}$. Clearly $\mathcal{C}$ is a cover of $\mathcal{L}$: $\cup_{\mathcal{C}} C = \mathcal{L}$.

A **lower bound** of $D \subseteq \mathcal{L}$ with respect to $\preceq$ is defined as $L_o(D) \in \mathcal{L}$ such that $\forall L \in D$, $L_o(D) \preceq L$. For a generic D, a lower bound need not exist, nor belong to D itself, nor be unique. In the case in which a lower bound exists and belongs to the set, it is necessarily unique, and one refers to it as the **smallest element** of D. Indicating with $\mathcal{C}_o$ the sub-collection of maximal chains admitting a lower bound, we have:

[2] In *regression settings*; that is, when the interest is focused on the dependence of one (or a set of) response(s) on a set of predictors, projective representations that entirely preserve information relative to the response can be identified using conditional independence. This is the object of a *theory of sufficient dimension reduction* which has grown through the contributions of several authors during the last decade. See for example R.D. Cook, 1998, F. Chiaromonte, R.D. Cook, 1997, Ker Chau Li, 1991, and references therein.

Proposition 1.1 *For any $C \in \mathcal{C}_o$, $L_o(C) \in C$ and it is unique.*

Proof: Suppose $\exists L_o(C)$. Since a maximal chain groups all comparable elements, $L_o(C) \preceq L$ for $L \in C$ automatically implies $L_o(C) \in C$; any lower bound of C must be contained in it. Now, take two lower bounds $L_{o,1}(C), L_{o,2}(C)$. By the above reasoning, they are both in C. Thus, one has both $L_{o,1}(C) \preceq L_{o,2}(C)$ and $L_{o,2}(C) \preceq L_{o,1}(C)$, which in turn implies $L_{o,1}(C) = L_{o,2}(C)$ by $P(3)$. □

A **minimal element** of $\preceq$ is defined as $L_o \in \mathcal{L}$ such that $L \preceq L_o$ implies $L = L_o$. We indicate with $\mathcal{L}_o$ the subset of minimal elements. As suggested by the intuition, minimal elements are all and only smallest elements of maximal chains which are bounded below:

Proposition 1.2 $\mathcal{L}_o = \{L_o(C)\ ,\ C \in \mathcal{C}_o\}$.

Proof: Suppose $\exists L_o(C)$ and let $L \preceq L_o(C)$. Since $L_o(C) \in C$ and a maximal chain groups all comparable elements, $L \in C$. Thus, one has also $L_o(C) \preceq L$, and $L = L_o(C)$ by $P(3)$. It follows that $L_o(C) \in \mathcal{L}_o$. Now, take $L_o \in \mathcal{L}_o$. Since $\mathcal{C}$ is a cover, there is at least one maximal chain containing it, say $L_o \in C$. Moreover, L_o being a minimal element, any $L \in C$ such that $L \preceq L_o$ must be $L = L_o$. So C is indeed bounded below by $L_o(C) = L_o \in C$. □

Notice that $\mathcal{L}_o \subseteq \cup_{\mathcal{C}_o} C$. We call **disjoint** (or triangulated) a partial order such that any two elements having a common predecessor or a common descendant are comparable:

- $P(4)$: if $\exists L$ such that $L \preceq L_1$ and $L \preceq L_2$, or $L_1 \preceq L$ and $L_2 \preceq L$, then $L_1 \preceq L_2$ or $L_2 \preceq L_1$ (lower and upper triangulation).

It is easy to show that lower and upper triangulation eliminate overlappings among maximal chains, and therefore that:

Proposition 1.3 *The collection $\mathcal{C}$ of maximal chains of a disjoint partial order is a partition of $\mathcal{L}$.*

Proof: Take $C_1, C_2 \in \mathcal{C}$, and let $L \in C_1 \cap C_2$. One has that $\forall L_1 \in C_1$, $L_1 \preceq L$ or $L \preceq L_1$, and $\forall L_2 \in C_2$, $L_2 \preceq L$ or $L \preceq L_2$. In any of the combinations, using $P(2)$ or $P(4)$ one obtains $L_1 \preceq L_2$ or $L_2 \preceq L_1$. Thus, $C_1 \cup C_2$ is a chain, which in turn implies $C_1 = C_2$, as both C_1 and C_2 are maximal. It follows that two distinct maximal chains $C_1 \neq C_2$ have $C_1 \cap C_2 = \emptyset$, and hence that the cover $\mathcal{C}$ is indeed a partition of $\mathcal{L}$. □

It is also easy to show that the partition constituted by maximal chains coincides with the one induced by the relation

$$L_1 \sim L_2 \ \text{ iff } \ L_1 \preceq L_2 \ \text{ or } \ L_2 \preceq L_1$$

which is indeed an equivalence relation under our assumptions [3].

Since maximal chains do not overlap, the (unique) smallest elements $L_o(C)$ of distinct C's $\in \mathcal{C}_o$ are distinct, and $\mathcal{L}_o = \{L_o(C)\ ,\ C \in \mathcal{C}_o\}$ indexes one-to-one-onto the sub-collection of maximal chains that are bounded below.

The advantage of specifying a definition/partition through a disjoint partial order resides exactly in the complete order it induces among laws with the same structure: If

[3] $\sim$ is reflexive by $P(1)$ and symmetric by construction. Under $P(2)$ and $P(4)$, $\sim$ is also transitive: Suppose $L_1 \sim L_2$ and $L_2 \sim L_3$. If $L_1 \preceq L_2$ and $L_2 \preceq L_3$, or $L_2 \preceq L_1$ and $L_3 \preceq L_2$, $L_1 \sim L_3$ follows from $P(2)$. If $L_2 \preceq L_1$ and $L_2 \preceq L_3$, or $L_1 \preceq L_2$ and $L_3 \preceq L_2$, $L_1 \sim L_3$ follows from $P(4)$.

the order corresponds to a simplification criterion, any law smaller (simpler) than L can be taken in lieu of L itself in order to investigate its structure. Moreover, for each $C \in \mathcal{C}_o$ there is a natural privileged representative for structural investigation; namely its (unique) smallest element. This is the *simplest version* of the structure shared by the laws in C. Privileged representatives of distinct C's $\in \mathcal{C}_o$ are distinct, and minimal elements can be taken as a *sub-repertoire of possible structures* according the definition/partition specified through $\preceq$ (the repertoire of all structures that are "bounded below"). Finally, one can define an operator $R : \cup_{\mathcal{C}_o} C \to \mathcal{L}_o$ which associates to each law in $\cup_{\mathcal{C}_o} C$ the smallest element of the maximal chain it belongs to:

$$R[L] = L_o(C) \ , \ \forall L \in C \ , \forall C \in \mathcal{C}_o$$

Upper triangulation guarantees that $R[L]$ is unique for each $L \in \cup_{\mathcal{C}_o} C$, and lower triangulation guarantees that $R[L_1] = R[L_2]$ iff L_1 and L_2 belong to the same $C \in \mathcal{C}_o$. Notice that minimal elements are all and only elements that are left unchanged by the operator: $R[L] = L$ iff $L \in \mathcal{L}_o$.

Now, consider a **bounded** disjoint partial order on $\mathcal{L}$; that is, a disjoint partial order satisfying

- $P(5)$: $\forall C \in \mathcal{C}$, $\exists L_o(C)$.

Since all maximal chains are bounded below ($\mathcal{C} = \mathcal{C}_o$), *each* $C \in \mathcal{C}$ has as privileged representative its (unique) smallest element. Privileged representatives of distinct C's $\in \mathcal{C}$ are distinct, and minimal elements can be taken as a *complete* repertoire of possible structures. Moreover, R is defined on the *whole* $\mathcal{L}$: for any $L \in \mathcal{L}$, $R[L]$ *can be taken in lieu of L itself to investigate its structure.* Consequently, the analysis can be restricted to the smallest linear subspace supporting $R[L]$, $\mathbf{S}(R[L])$, without loss of structural information. If $\dim(\mathbf{S}(R[L])) < k$, we refer to this as **exhaustive dimension reduction**.

When using a bounded and disjoint partial order to generate a definition/partition, the complete order among laws with the same structure becomes instrumental, as it is always possible to refer directly to the smallest element. The focus is therefore on the repertoire of minimal elements, and on the operator associating laws to them. We call the laws in $\mathcal{L}_o$ **origins** of $\preceq$, and $R : \mathcal{L} \to \mathcal{L}_o$ **reduction operator** of $\preceq$. Since origins are all and only fixed points of the reduction operator, we also refer to them as **irreducible laws**.

1.1 Invariances

We call **invariances** of $\preceq$ invertible transformations that commute with reduction

$$\mathcal{T} = \{ \text{ invertible } T : \mathcal{L} \to \mathcal{L} \text{ s.t. } RT[L] = TR[L], \forall L \in \mathcal{L}\}$$

Clearly the identity $I \in \mathcal{T}$, and therefore $\mathcal{T} \neq \emptyset$. Moreover, $\mathcal{T}$ is a *group with composition*. In fact, $\mathcal{T}$ is closed with respect to composition: $\forall T_1, T_2 \in \mathcal{T}$, $\forall L \in \mathcal{L}$, $T_1T_2R[L] = T_1RT_2[L] = RT_1T_2[L]$. Thus, it inherits the group structure from the class of all invertible transformations. It is easy to show that invariances are all and only invertible transformations that preserve both irreducibility and equivalence:

Proposition 1.4 *$T \in \mathcal{T}$ if and only if $L_o \in \mathcal{L}_o \Rightarrow T[L_o] \in \mathcal{L}_o$ and $L_1 \sim L_2 \Rightarrow T[L_1] \sim T[L_2]$.*

Proof: Suppose $RT[L] = TR[L], \forall L \in \mathcal{L}$. Take $L_o \in \mathcal{L}_o$, i.e. such that $L_o = R[L_o]$. Then $T[L_o] = TR[L_o] = RT[L_o]$, i.e. $T[L_o] \in \mathcal{L}_o$. Now take $L_1 \sim L_2$, i.e. such that

$R[L_1] = R[L_2]$. Then $TR[L_1] = TR[L_2]$ and thus $RT[L_1] = RT[L_2]$, i.e. $T[L_1] \sim T[L_2]$. For the opposite implication, consider $L \in \mathcal{L}$. Since $L_o \in \mathcal{L}_o \Rightarrow T[L_o] \in \mathcal{L}_o$, one has $TR[L] \in \mathcal{L}_o$. Moreover, since $L_1 \sim L_2 \Rightarrow T[L_1] \sim T[L_2]$, $L \sim R[L]$ implies $T[L] \sim TR[L]$. One can conclude that $RT[L] = TR[L]$. □

Notice that as $T^{-1} \in \mathcal{T}$, $\forall T \in \mathcal{T}$, one also has $T[L_o] \in \mathcal{L}_o \Rightarrow L_o \in \mathcal{L}_o$ and $T[L_1] \sim T[L_2] \Rightarrow L_1 \sim L_2$. Notice also that a $T \in \mathcal{T}$ does not necessarily preserve the ordering within each C –except for the "bottom"; that is, transforming an origin into another origin. Although this is irrelevant here (the ordering within each class is instrumental to the identification of privileged members and the definition of the reduction operator), any $\preceq$-preserving invertible transformation is clearly an invariance according to our definition. In particular, the class of all $\preceq$-preserving invertible transformations is a sub-group of $\mathcal{T}$.

Since for an invariance $R[L] = T^{-1}TR[L]$ coincides with $T^{-1}RT[L]$, $RT[L]$ *can be taken in lieu of* L *to investigate its structure.* The advantage is that *reduction on the* T*-scale might be easier to perform* than that on the original one.

1.2 Invariant marginalizations and kernels

Let $S \subseteq \mathbb{R}^k$ be a linear subspace, and $\mathcal{L}_S$ the collection of all probability distributions supported by S. The marginalization to S, $M_S : \mathcal{L} \to \mathcal{L}_S$, is defined as

$$M_S[L](B) = \int_{\{x: P_S x \in B\}} L(dx) \quad , \quad \forall \text{ (meas.) } B \subseteq S$$

($P_{(\cdot)}$ is the orthogonal projector operator on the argument subspace with respect to the standard inner product in $\mathbb{R}^k$). Clearly, the elements of $\mathcal{L}_S$ are all and only laws which are left unchanged by M_S: $M_S[L] = L$ iff $L \in \mathcal{L}_S$.

After an appropriate isomorphism, the whole framework we have set up for $\mathbb{R}^k$ can be reproduced *within* S [4]. Notation-wise, the various components of the framework within the subspace will be sub-indexed by S.

Consider an origin $L_o \in \mathcal{L}_o$. We call **invariant marginalizations** of $\preceq$ **on** L_o the marginalizations that commute with reduction *when restricted to laws equivalent to* L_o (i.e. to the class of which L_o is the privileged member)

$$\begin{aligned} \mathcal{M}(L_o) \quad = \quad & \{M_S : \mathcal{L} \to \mathcal{L}_S \text{ s.t. } R_S M_S[L] = \\ & M_S R[L] = M_S[L_o], \forall L \sim L_o\} \end{aligned}$$

Clearly $I \in \mathcal{M}(L_o)$ and therefore $\mathcal{M}(L_o) \neq \emptyset$. The invariant marginalizations on L_o are all and only marginalizations that preserve irreducibility of L_o and equivalence to L_o:

Proposition 1.5 *Let* $L_o \in \mathcal{L}_o$. $M_S \in \mathcal{M}(L_o)$ *if and only if* $M_S[L_o] \in \mathcal{L}_{o,S}$ *and* $L \sim L_o \Rightarrow M_S[L] \sim_S M_S[L_o]$.

Proof: Suppose $R_S M_S[L] = M_S[L_o]$, $\forall L \sim L_o$. $L_o = R[L_o]$ gives $M_S[L_o] = M_S R[L_o] = R_S M_S[L_o]$, that is $M_S[L_o] \in \mathcal{L}_{o,S}$. Now take an $L \sim L_o$, i.e. such that $R[L] = L_o$. Then $M_S R[L] = M_S[L_o]$ and thus $R_S M_S[L] = M_S[L_o]$, i.e. $M_S[L] \sim_S M_S[L_o]$. For the opposite

[4] This statement needs further qualification if the partial order one is using depends on the choice of coordinate system (i.e. of orthonormal basis for the Euclidean space), or cannot be reproduced in an arbitrary dimension $< k$. In particular, for the former case, one would be linked to particular bases for both $\mathbb{R}^k$ and $S \leftrightarrow \mathbb{R}^{dim(S)}$. Notice that no problem would arise if, in whatever dimension, *rotations* (i.e. changes of orthonormal basis) were invariances, and thus, a fortiori, if they preserved the ordering. Some details on rotations are given in Section 3.

implication, consider again $L \sim L_o$. One has $M_S[L_o] \in \mathcal{L}_{o,S}$. Moreover, $L \sim L_o \Rightarrow M_S[L] \sim_S M_S[L_o]$. It follows that $R_S M_S[L] = M_S[L_o]$. $\square$

Once more, a $M_S \in \mathcal{M}(L_o)$ is not required to preserve the ordering in the class of L_o, or any other.

We call **kernels** of $\preceq$ those origins on which all marginalizations are invariant

$$\mathcal{K} = \{L_o \in \mathcal{L}_o \text{ s.t. } M_S \in \mathcal{M}(L_o) \ , \ \forall S \subseteq \mathbb{R}^k\}$$

Whether $\mathcal{K} \neq \emptyset$, and how to further characterize kernels, will depend on the partial order under consideration.

If $R[L] \in \mathcal{K}$, there are two immediate advantages. First, since $R_S M_S[L] = M_S R[L]$, $\forall S \subseteq \mathbb{R}^k$, any marginal of $R[L]$ can be investigated by reducing within S the marginal of L to S. Second, *the meaning of exhaustive dimension reduction is stronger.* In fact, not only one can restrict the analysis to $\mathbf{S}(R[L])$ without loss of structural information, but

$$R[L] = M_{\mathbf{S}(R[L])} R[L] = R_{\mathbf{S}(R[L])} M_{\mathbf{S}(R[L])}[L]$$

so that reducing within $\mathbf{S}(R[L])$ the marginal of L to $\mathbf{S}(R[L])$ is exactly equivalent to reducing L in $\mathbb{R}^k$. If indeed $\dim(\mathbf{S}(R[L])) < k$, *the reduction in lower dimension might be easier to perform.*

2 Generating partial orders via binary operations and noise classes

Under proper assumptions, a bounded and disjoint partial order can be generated via a **binary operation** and a **noise class**. Let $\circ : \mathcal{L} \times \mathcal{L} \to \mathcal{L}$ be a binary operation such that:

- $OP(1)$: $\forall L_1, L_2, L_3 \in \mathcal{L}$, $L_1 \circ (L_2 \circ L_3) = (L_1 \circ L_2) \circ L_3$ (associativity)
- $OP(2)$: $\exists$ (unique) $0 \in \mathcal{L}$ such that $\forall L \in \mathcal{L}$, $L \circ 0 = 0 \circ L = L$
- $OP(3)$: $\forall L \in \mathcal{L}$, $L_1 \circ L = L_2 \circ L$ or $L \circ L_1 = L \circ L_2$ implies $L_1 = L_2$ (cancelability) [5].

Also, let $\mathcal{E} \subseteq \mathcal{L}$ be a class which, relative to $\circ$, satisfies:

- $ER(1)$: $0 \in \mathcal{E}$
- $ER(2)$: $\forall \varepsilon_1, \varepsilon_2 \in \mathcal{E}$, $\varepsilon_1 \circ \varepsilon_2 \in \mathcal{E}$ (closure)
- $ER(3)$: $\forall L \in \mathcal{L}$, $\forall \varepsilon \in \mathcal{E} \setminus \{0\}$, $0 \neq L \circ \varepsilon$ [6]

[5] $L_1 = L_2$ always implies $L_1 \circ L = L_2 \circ L$ and $L \circ L_1 = L \circ L_2$; one could label L (right and left) *discriminant element with respect to* $\circ$ if one of the latter is enough to imply the former. If all $\in \mathcal{L}$ are discriminant elements, then $\circ$ is called *cancelable.* Clearly 0, as well as any element of $\mathcal{L}$ endowed with a (right and left) inverse with respect to $\circ$, is a discriminant element. Cancelability is not as strong as requiring the existence of an inverse for all elements, though. As an example, consider $\mathbb{N}$ with $+$: no element except 0 has an inverse, and yet every $n \in \mathbb{N}$ is such that $n_1 + n = n_2 + n$ or $n + n_1 = n + n_2$ is enough to conclude $n_1 = n_2$.

[6] That is, no element of $\mathcal{E}$, except 0, admits a (left) inverse with respect to $\circ$.

- $ER(4)$: $\forall \varepsilon_1, \varepsilon_2 \in \mathcal{E}$, $\exists \tilde{\varepsilon}, \breve{\varepsilon} \in \mathcal{E}$ such that $\varepsilon_2 = \varepsilon_1 \circ \tilde{\varepsilon}$ or $\varepsilon_1 = \varepsilon_2 \circ \tilde{\varepsilon}$ and $\varepsilon_2 = \breve{\varepsilon} \circ \varepsilon_1$ or $\varepsilon_1 = \breve{\varepsilon} \circ \varepsilon_2$ (triangulations) [7]
- $ER(5)$: $\forall L \in \mathcal{L}$, $\exists \varepsilon \in \mathcal{E}$ such that (1) $L = \tilde{L} \circ \varepsilon$ for some $\tilde{L} \in \mathcal{L}$, and (2) $\forall \check{L} \in \mathcal{L}$, $\forall \breve{\varepsilon} \in \mathcal{E} \setminus \{0\}$, $L \neq \check{L} \circ (\breve{\varepsilon} \circ \varepsilon)$.

Notice that $ER(3)$ is a special case of $ER(5)$. In terms of the partial order we are about to introduce, $ER(5)$ requires existence of a largest element for each set

$$\mathcal{E}(L) = \{\varepsilon \in \mathcal{E} \text{ s.t. } L = \tilde{L} \circ \varepsilon \text{ for some } \tilde{L} \in \mathcal{L}\}$$

which, as we will see, is a means to impose boundedness. Define

$$L_1 \preceq L_2 \quad \text{iff} \quad \exists \varepsilon \in \mathcal{E} \ \text{ such that } \ L_2 = L_1 \circ \varepsilon$$

With the above assumptions, $\preceq$ is a bounded disjoint partial order on $\mathcal{L}$:

Proposition 2.1 *Under $OP(1)$, $OP(2)$, $OP(3)$ and $ER(1)$, $ER(2)$, $ER(3)$, $ER(4)$, $\preceq$ satisfies $P(1)$, $P(2)$, $P(3)$, $P(4)$. If in addition $ER(5)$ holds, $\preceq$ satisfies also $P(5)$.*

Proof: $P(1)$ holds by $OP(2)$ and $ER(1)$: $\forall L \in \mathcal{L}$, $L = L \circ 0$, and $0 \in \mathcal{E}$.

$P(2)$ becomes: If $\exists \varepsilon_1, \varepsilon_2 \in \mathcal{E}$ such that $L_2 = L_1 \circ \varepsilon_1$ and $L_3 = L_2 \circ \varepsilon_2$, then $\exists \varepsilon \in \mathcal{E}$ such that $L_3 = L_1 \circ \varepsilon$. In fact $L_3 = L_2 \circ \varepsilon_2 = (L_1 \circ \varepsilon_1) \circ \varepsilon_2 = L_1 \circ (\varepsilon_1 \circ \varepsilon_2)$ by $OP(1)$, with $\varepsilon_1 \circ \varepsilon_2 \in \mathcal{E}$ by $ER(2)$.

$P(3)$ becomes: If $\exists \varepsilon_1, \varepsilon_2 \in \mathcal{E}$ such that $L_2 = L_1 \circ \varepsilon_1$ and $L_1 = L_2 \circ \varepsilon_2$, then $L_1 = L_2$. In fact $L_1 \circ 0 = L_1$ by $OP(2)$, which is $= L_2 \circ \varepsilon_2 = (L_1 \circ \varepsilon_1) \circ \varepsilon_2 = L_1 \circ (\varepsilon_1 \circ \varepsilon_2)$ by $OP(1)$, which implies $0 = \varepsilon_1 \circ \varepsilon_2$ by $OP(3)$. This in turn implies $\varepsilon_2 = 0$ by $ER(3)$, and therefore $L_1 = L_2$ by $OP(2)$.

$P(4)$ becomes: If $\exists L$ and $\varepsilon_1, \varepsilon_2 \in \mathcal{E}$ such that $L_1 = L \circ \varepsilon_1$ and $L_2 = L \circ \varepsilon_2$, or $L = L_1 \circ \varepsilon_1$ and $L = L_2 \circ \varepsilon_2$, then $\exists \varepsilon \in \mathcal{E}$ such that $L_2 = L_1 \circ \varepsilon$ or $L_1 = L_2 \circ \varepsilon$. Suppose one had $L_1 = L \circ \varepsilon_1$ and $L_2 = L \circ \varepsilon_2$, and for instance ($ER(4)$) $\varepsilon_2 = \varepsilon_1 \circ \tilde{\varepsilon}$. Then $L_2 = L \circ \varepsilon_2 = L \circ (\varepsilon_1 \circ \tilde{\varepsilon}) = (L \circ \varepsilon_1) \circ \tilde{\varepsilon}$ by $OP(1)$, which is $= L_1 \circ \tilde{\varepsilon}$ (analogously, $\varepsilon_1 = \varepsilon_2 \circ \tilde{\varepsilon}$ would lead to $L_1 = L_2 \circ \tilde{\varepsilon}$). On the other hand, suppose $L = L_1 \circ \varepsilon_1$ and $L = L_2 \circ \varepsilon_2$, and for instance ($ER(4)$ again) $\varepsilon_1 = \breve{\varepsilon} \circ \varepsilon_2$. Then $L_2 \circ \varepsilon_2 = L_1 \circ \varepsilon_1 = L_1 \circ (\breve{\varepsilon} \circ \varepsilon_2) = (L_1 \circ \breve{\varepsilon}) \circ \varepsilon_2$ by $OP(1)$, which implies $L_2 = L_1 \circ \breve{\varepsilon}$ by $OP(3)$ (analogously, $\varepsilon_2 = \breve{\varepsilon} \circ \varepsilon_1$ would lead to $L_1 = L_2 \circ \breve{\varepsilon}$).

Finally, $P(5)$ holds under $ER(5)$: Take any maximal chain $C \in \mathcal{C}$, fix an $L \in C$, and let $\tilde{L}$ be the corresponding law as from $ER(5.1)$. Clearly $\tilde{L} \in C$, so any $\check{L} \in C$ is comparable to it; that is, $\exists \breve{\varepsilon} \in \mathcal{E}$ such that $\tilde{L} = \check{L} \circ \breve{\varepsilon}$ or $\check{L} = \tilde{L} \circ \breve{\varepsilon}$. If the former were the case, one would have $L = (\check{L} \circ \breve{\varepsilon}) \circ \varepsilon = \check{L} \circ (\breve{\varepsilon} \circ \varepsilon)$ by $ER(5.1)$ and $OP(2)$, which contradicts $ER(5.2)$ unless $\breve{\varepsilon} = 0$ (i.e. $\check{L} = \tilde{L}$ itself. It follows that $\tilde{L} \in C$ bounds C from below [8]. $\square$

In summary, under proper assumptions the couple $(\circ, \mathcal{E})$ can be used to generate an appropriate partial order $\preceq$, and hence a definition/partition $\mathcal{C}$. The operation can be interpreted as a *superposition mechanism*, and the noises as *no-structure laws*. Because of $ER(4)$, the noise class is one of the equivalence classes: $\mathcal{E} \in \mathcal{C}$. As a consequence, it is completely ordered by $\preceq$. One has

$$\varepsilon_1 \preceq \varepsilon_2 \ \leftrightarrow \varepsilon_2 = \varepsilon_1 \circ \varepsilon$$

[7] Notice $ER(4)$ is implied by $ER(4.a)$: $\forall \varepsilon_1, \varepsilon_2 \in \mathcal{E}$, $\exists \varepsilon \in \mathcal{E}$ such that $\varepsilon_2 = \varepsilon_1 \circ \varepsilon$ or $\varepsilon_1 = \varepsilon_2 \circ \varepsilon$, and $ER(4.b)$: $\forall \varepsilon_1, \varepsilon_2 \in \mathcal{E}$, $\varepsilon_1 \circ \varepsilon_2 = \varepsilon_2 \circ \varepsilon_1$ (commutativity within $\mathcal{E}$).

[8] $\tilde{L}$ is indeed the (unique) smallest element of the maximal chain. Notice this makes all the $\tilde{L}$'s associated by $ER(5.1)$ to the various laws in C coincide with $L_o(C)$.

Moreover, because of $ER(1)$ and $ER(3)$, the origin of the noise class is the null element of the binary operation: $L_o(\mathcal{E}) = 0$. In parallel to the ordering in $\mathcal{E}$, the ordering within each equivalence class corresponds to containing larger or smaller noise terms. For $L_1, L_2 \in C$ one has

$$\begin{aligned} L_1 \preceq L_2 \quad &\leftrightarrow \quad L_2 = L_1 \circ \varepsilon \\ &\leftrightarrow \quad L_1 = L_o(C) \circ \varepsilon_1 \ , \ L_2 = L_o(C) \circ \varepsilon_2 \\ &\qquad\ \ \text{with } \ \varepsilon_1 \preceq \varepsilon_2 \end{aligned}$$

The origin $L_o(C)$ of each class is the one and only member containing 0-noise; that is, the one and only element of the class in which no noise is superimposed to the structure to be observed. Thus, using $R[L]$ in lieu of L to investigate its structure acquires a precise simplification and clarification meaning: we are eliminating from the analysis as large a noise as possible, to remain with the skeleton structure of L. Notice that here, as in Factor Analysis, we have a *decomposition model* which separates a term exhaustively embodying the structure from a noise that does not contribute to it, and can therefore be neglected.

One advantage of specifying a partial order, and hence a definition/partition, through appropriate binary operation and noise class, is the interpretations it provides for origins and reduction operator. A second advantage is the possibility of further characterizing invariances, invariant marginalizations, origins and kernels via sufficient conditions.

2.1 Some sufficient conditions

Let us start with **sufficient conditions for preserving equivalence**. Consider a transformation $T : \mathcal{L} \to \mathcal{L}$.

Proposition 2.2 *If* $\exists U_T : \mathcal{L} \to \mathcal{L}$ *such that* (1) $\varepsilon \in \mathcal{E} \Rightarrow U_T[\varepsilon] \in \mathcal{E}$, *and* (2) $\forall L_o \in \mathcal{L}_o$, $\forall \varepsilon \in \mathcal{E}$, $T[L_o \circ \varepsilon] = T[L_o] \circ U_T[\varepsilon]$, *then* $L_1 \sim L_2 \Rightarrow T[L_1] \sim T[L_2]$.

Proof: $L_1 \sim L_2$ translates into: $\exists \varepsilon_1, \varepsilon_2 \in \mathcal{E}$ and $L_o \in \mathcal{L}_o$ such that $L_1 = L_0 \circ \varepsilon_1$ and $L_2 = L_o \circ \varepsilon_2$. Thus, one has $T[L_1] = T[L_o \circ \varepsilon_1] = T[L_o] \circ U_T[\varepsilon_1]$ and $T[L_2] = T[L_o \circ \varepsilon_2]$ $= T[L_o] \circ U_T[\varepsilon_2]$ by (2), with $U_T[\varepsilon_1], U_T[\varepsilon_1] \in \mathcal{E}$ by (1). It follows that $T[L_1] \sim T[L_2]$. □

For an invertible T which admits such a U_T, being an invariance reduces to preserving irreducibility: $T \in \mathcal{T}$ if and only if $L_o \in \mathcal{L}_o \Rightarrow T[L_o] \in \mathcal{L}_o$. It is also easy to show that the sub-class of invariances identified by this proposition is closed with respect to composition, and therefore a sub-group of $\mathcal{T}$.

Now, consider a marginalization $M_S : \mathcal{L} \to \mathcal{L}_S$, and indicate with $\circ_S$ and $\mathcal{E}_S$ the binary operation and error class defined within the subspace S. We have that

Proposition 2.3 *Let* $L_o \in \mathcal{L}_o$. *If* $\exists U_S : \mathcal{L} \to \mathcal{L}_S$ *such that* (1) $\varepsilon \in \mathcal{E} \Rightarrow U_S[\varepsilon] \in \mathcal{E}_S$, *and* (2) $\forall \varepsilon \in \mathcal{E}$, $M_S[L_o \circ \varepsilon] = M_S[L_o] \circ_S U_S[\varepsilon]$, *then* $L \sim L_o \Rightarrow M_S[L] \sim_S M_S[L_o]$.

Proof: $L \sim L_o$ translates into: $\exists \varepsilon \in \mathcal{E}$ such that $L = L_o \circ \varepsilon$. So one has $M_S[L] = M_S[L_o \circ \varepsilon]$ $= M_S[L_o] \circ U_S[\varepsilon]$ by (2), with $U_S[\varepsilon] \in \mathcal{E}_S$ by (1). It follows that $M_S[L] \sim_S M_S[L_o]$. □

So for an M_S admitting such a U_S invariance on L_o reduces to preserving irreducibility of L_o itself: $M_S \in \mathcal{M}(L_o)$ if and only if $M_S[L_o] \in \mathcal{L}_{o,S}$.

Passing to **sufficient conditions for irreducibility**, let us indicate with $\mathrm{Cs}(L) \subseteq \mathbb{R}^k$ the closed support of a law, and consider the further assumption

- $OP(4)$: $\forall L_1, L_2 \in \mathcal{L}$, $\mathrm{Cs}(L_1 \circ L_2) \supseteq \mathrm{Cs}(L_i)$, $i = 1, 2$

Proposition 2.4 *Let $L \in \mathcal{L}$. Under $OP(4)$, if $Cs(L) \not\supseteq \bigcap_{\varepsilon\in\mathcal{E}\setminus\{0\}} Cs(\varepsilon)$, then $L \in \mathcal{L}_o$.*

Proof: L not being an origin translates into: $\exists \tilde{L} \in \mathcal{L}\setminus\{L\}$ and $\tilde{\varepsilon} \in \mathcal{E}\setminus\{0\}$ such that $L = \tilde{L} \circ \tilde{\varepsilon}$. But then $\mathrm{Cs}(L) = \mathrm{Cs}(\tilde{L}\circ\tilde{\varepsilon}) \supseteq \mathrm{Cs}(\tilde{\varepsilon}) \supseteq \bigcap_{\varepsilon\in\mathcal{E}\setminus\{0\}} \mathrm{Cs}(\varepsilon)$ by $OP(4)$. Hence, $\mathrm{Cs}(L) \not\supseteq \bigcap_{\varepsilon\in\mathcal{E}\setminus\{0\}} \mathrm{Cs}(\varepsilon)$ implies $L \in \mathcal{L}_o$. □

In particular, under $OP(4)$, $\mathrm{Cs}(L) \subset \bigcap_{\varepsilon\in\mathcal{E}\setminus\{0\}} \mathrm{Cs}(\varepsilon)$ implies $L \in \mathcal{L}_o$. Recall that the sub-class of origins identified by this proposition can be interpreted as a *sub-repertoire of possible structures.*

Last, regarding **kernels** we have that:

Proposition 2.5 *Let $L_o \in \mathcal{L}_o$. Under $OP(4)$, if for any subspace $S \subseteq \mathbb{R}^k$*
(1) $\exists U_S : \mathcal{L} \to \mathcal{L}_S$ such that (1.1) $\varepsilon \in \mathcal{E} \Rightarrow U_S[\varepsilon] \in \mathcal{E}_S$, and (1.2) $\forall \varepsilon \in \mathcal{E}$, $M_S[L_o \circ \varepsilon] = M_S[L_o] \circ_S U_S[\varepsilon]$
(2) $Cs(M_S[L_o]) \not\supseteq \bigcap_{\varepsilon\in\mathcal{E}_S\setminus\{0\}} Cs(\varepsilon)$
then $L_o \in \mathcal{K}$.

Proof: By Proposition 2.3, (1) guarantees that M_S preserves equivalence to L_o for any S. Furthermore, by Proposition 2.4, (2) guarantees that $M_S[L_o] \in \mathcal{L}_{o,S}$. Hence, $L_o \in \mathcal{K}$. □

3 Convolution and white noises; a first example

Take the binary operation to be convolution: $\circ = *$. Clearly, $*$ satisfies $OP(1)$, $OP(2)$ and $OP(3)$. The null element of $*$ is δ_{0_k}; the point-mass at $0_k \in \mathbb{R}^k$. Take as noise class the white noises of $\mathbb{R}^k$

$$\mathcal{E} = \mathcal{W} = \{N(0_k, \beta I_k), \beta \in \mathbb{R}^1_+\}$$

It is immediate to show that $\mathcal{W}$ satisfies $ER(1)$, $ER(2)$, $ER(3)$ and $ER(4)$ [9]. Hence, $(*, \mathcal{W})$ allow us to define a *disjoint partial order on $\mathcal{L}$*

$$L_1 \preceq L_2 \quad \text{iff} \quad \exists \beta \in \mathbb{R}^1_+ \ \text{ such that } \ L_2 = L_1 * N(0_k, \beta I_k)$$

and thereby a definition/partition $\mathcal{C}$, corresponding to the equivalence relation

$$\begin{aligned} L_1 \sim L_2 \quad \text{iff} \quad & \exists \beta \in \mathbb{R}^1_+ \ \text{ such that} \\ & L_2 = L_1 * N(0_k, \beta I_k) \ \text{ or} \\ & L_1 = L_2 * N(0_k, \beta I_k) \end{aligned}$$

The complete order within $\mathcal{W} \in \mathcal{C}$, and any other class, corresponds to the complete order in the one-dimensional parameterization of white noises ($\beta \in \mathbb{R}^1_+$).

Notice that $L_2 = L_1 * N(0_k, \beta I_k)$ can be rewritten in terms of characteristic functions as $\phi_{L_2}(u) = \phi_{L_1}(u)\ e^{-\frac{\beta}{2}\|u\|^2}$. For any $L \in \mathcal{L}$, let

$$\mathcal{B}(L) = \{\beta \in \mathbb{R}^1_+ \text{ s.t. } \phi_L(u)\ e^{\frac{\beta}{2}\|u\|^2} \text{ is a ch. fct. on } \mathbb{R}^k\}$$

This set is in one-to-one correspondence, via characteristic functions, with the predecessors of L. It is certainly non-empty, as it must contain 0. Moreover, $\mathcal{B}(L)$ *is a convex, closed*

[9] In particular, it is enough to observe one of the triangulations, as the other will follow immediately from the fact that $*$ is commutative on the whole $\mathcal{L}$, and therefore a fortiori on $\mathcal{W}$.

and bounded subset of $\mathbb{R}^1_+$. Let $\beta_1, \beta_2 \in \mathcal{B}(L)$, $0 \leq \beta_1 \leq \beta_2$, and $\lambda \in [0,1]$. Define $\beta_\lambda = \lambda\beta_1 + (1-\lambda)\beta_2$, and consider

$$\phi_L(u)\ e^{\frac{\beta_\lambda}{2}\|u\|^2} = \left[\phi_L(u)\ e^{\frac{\beta_2}{2}\|u\|^2}\right]\left[\ e^{-\frac{(\beta_2-\beta_\lambda)}{2}\|u\|^2}\right]$$

The first factor is a characteristic function, as $\beta_2 \in \mathcal{B}(L)$. The second factor is the characteristic function of a $N(0_k, (\beta_2 - \beta_\lambda)I_k))$, as $(\beta_2 - \beta_\lambda) \geq 0$. Thus, the product is a characteristic function, and $\beta_\lambda \in \mathcal{B}(L)$. Next, consider a converging sequence $\beta_n \in \mathcal{B}(L)$, $n = 1, 2 \ldots$, $\beta_n \to \beta \geq 0$. For any $u \in \mathbb{R}^k$, $e^{\frac{(\cdot)}{2}\|u\|^2}$ is continuous, so

$$\phi_L(u)\ e^{\frac{\beta}{2}\|u\|^2} = \lim_{n\to\infty} \phi_L(u)\ e^{\frac{\beta_n}{2}\|u\|^2}$$

which is a limit of characteristic functions, as $\beta_n \in \mathcal{B}(L)$, $n = 1, 2 \ldots$. Furthermore, it is continuous in $u = 0_k$, as both $\phi_L(\cdot)$ and $e^{\frac{\beta}{2}\|(\cdot)\|^2}$ are. Thus, it is itself a characteristic function, and $\beta \in \mathcal{B}(L)$. Last, fix an amount of mass strictly smaller than 1, for example .5. There exists a finite radius $0 \leq \rho < +\infty$ such that the closed ball $C_\rho = \{x \in \mathbb{R}^k \text{ s.t. } \|x\| \leq \rho\}$ has $L(C_\rho) > .5$. Correspondingly, there exists a finite variance $0 \leq \bar\beta < +\infty$ such that $\forall\tilde\beta \in \mathbb{R}^1_+ \setminus \{0\}, \forall x \in \mathbb{R}^k$ one has

$$N(0_k, (\bar\beta + \tilde\beta)I_k)(C_\rho - x) \leq N(0_k, (\bar\beta + \tilde\beta)I_k)(C_\rho) < .5$$

It follows that $\forall\tilde L \in \mathcal{L}, \forall\tilde\beta \in \mathbb{R}^1_+ \setminus \{0\}$ one has

$$\begin{aligned}(\tilde L * N(0_k, (\bar\beta + \tilde\beta)I_k))(C_\rho) &= \\ \textstyle\int_{\mathbb{R}^k} N(0_k, (\bar\beta + \tilde\beta)I_k)(C_\rho - x)\tilde L(dx) &< .5\end{aligned}$$

and therefore that

$$\forall\tilde L \in \mathcal{L}, \forall\tilde\beta \in \mathbb{R}^1_+ \setminus \{0\} \quad , \quad L \neq \tilde L * N(0_k, (\bar\beta + \tilde\beta)I_k)$$

so any $\beta \in \mathcal{B}(L)$ must be $\beta \leq \bar\beta < +\infty$. In summary, we can write $\mathcal{B}(L) = [0, \beta_o(L)]$ with $\beta_o(L) \leq \bar\beta < +\infty$.

It is immediate to see that *origins are all and only those laws for which* $\mathcal{B}(L)$ *is trivial*:

$$\mathcal{L}_o = \{L \in \mathcal{L} \text{ s.t. } \beta_o(L) = 0\}$$

and that *the reduction operator associates to* L *the law identified by the characteristic function*

$$\phi_{R[L]}(u) = \phi_L(u)\ e^{\frac{\beta_o(L)}{2}\|u\|^2}$$

Notice that closure and boundedness guarantee that $\mathcal{B}(L)$ has a (unique) largest element, $\beta_o(L)$, for any $L \in \mathcal{L}$. This corresponds exactly to $ER(5)$: *Our disjoint partial order is bounded.* It follows that origins represent a complete repertoire of structures, and R is defined on the whole $\mathcal{L}$.

If L admits finite second order moments, the smallest supporting linear subspace is given by

$$\mathbf{S}(L) = \mathrm{Sp}(\mathrm{Cov}(L)) \oplus \mathrm{Sp}(\mathrm{E}(L))$$

where $\oplus$ is the direct sum between subspaces, and $\mathrm{Sp}(\cdots)$ the span of the argument vectors –or column vectors of the argument matrix (see for example M.L. Eaton, 1983 and F. Chiaromonte, 1997). We clearly have

$$\mathrm{E}(R[L]) = \mathrm{E}(L) \quad , \quad \mathrm{Cov}(R[L]) = \mathrm{Cov}(L) - \beta_o(L)I_k$$

Rewriting $\mathrm{Cov}(L) = \Gamma(L)'\mathrm{Di}(\eta_i(L))\Gamma(L)$, where $\eta_i(L)$, $i = 1, \ldots k$ are the eigenvalues in non-increasing order, $\mathrm{Di}(\eta_i(L))$ is diagonal, and $\Gamma(L)$ is the rotation to a spectral basis of $\mathrm{Cov}(L)$ [10], we have

$$\mathrm{Cov}(R[L]) = \Gamma(L)'\mathrm{Di}(\eta_i(L) - \beta_o(L))\Gamma(L)$$

with the same rotation, and eigenvalues all decreased by $\beta_o(L)$. Non-negative definiteness of covariance operators requires then $\beta_o(L) \leq \eta_k(L)$, the smallest eigenvalue of $\mathrm{Cov}(L)$. Moreover, indicating with $\nu(L)$ the multiplicity of the smallest eigenvalue, with $\Gamma_i(L)$, $i = 1, \ldots k$ the eigenvectors, and with $\mathrm{In}(\cdot)$ the indicator function of the argument condition, one has (F. Chiaromonte, 1997)

$$\begin{aligned} \mathbf{S}(R[L]) &= \mathrm{Sp}(\Gamma_1(L), \ldots, \Gamma_{\nu(L)-1}(L)) \\ &\oplus \mathrm{In}(\beta_o(L) < \eta_k(L))\ \mathrm{Sp}(\Gamma_{\nu(L)}, \ldots, \Gamma_k) \\ &\oplus \mathrm{Sp}(\mathrm{E}(L)) \end{aligned}$$

Convolution also satisfies $OP(4)$, and

$$\bigcap_{\omega \in \mathcal{W} \setminus \{\delta_{0_k}\}} \mathrm{Cs}(\omega) = \mathbb{R}^k$$

Thus, due to Proposition 2.4, *all laws whose closed support does not cover the entire* $\mathbb{R}^k$ *are irreducible*. Those include: Laws with "thick" holes with respect to the Lebesgue measure on $\mathbb{R}^k$, i.e. for which $\exists$ (measurable) B with $Leb(B) > 0$, but $L(B) = 0$; laws entirely supported by affine subspaces of $\mathbb{R}^k$, i.e. whose affine support is $\mathrm{As}(L) \subset \mathbb{R}^k$; laws whose closed support $\mathrm{Cs}(L)$ (whether full-dimensional or not) is bounded, etc. There are irreducible laws such that $\mathrm{Cs}(L_o) = \mathbb{R}^k$, though. For example, laws with a density with respect to the Lebesgue measure on $\mathbb{R}^k$, whose tails vanish faster than that of any non-degenerate $N(0_k, \beta I_k)$ in at least some directions (F. Chiaromonte, 1996).

Let $\Gamma \in \mathcal{G}$ be a rotation of $\mathbb{R}^k$, $v \in \mathbb{R}^k$, $\gamma \in \mathbb{R}^1$, and define an affine rotation $G_{v,\gamma,\Gamma} : \mathcal{L} \to \mathcal{L}$ as

$$G_{v,\gamma,\Gamma}[L](B) = \int_{\{x: v+\gamma\Gamma x \in B\}} L(dx) \quad , \quad \forall \text{ (meas.) } B \subseteq \mathbb{R}^k$$

It is easy to show that $\mathcal{W}$ is closed under, and only under, centered affine rotations; that is

$$\omega \in \mathcal{W} \Rightarrow U[\omega] \in \mathcal{W} \quad \text{iff} \quad \exists \gamma \in \mathbb{R}^1_+, \Gamma \in \mathcal{G} \ \text{ such that } \ U = G_{0_k,\gamma,\Gamma}$$

(see for example M.L. Eaton, 1983). Consequently, Proposition 2.2 can be rephrased as: A sufficient condition for T to give $L_1 \sim L_2 \Rightarrow T[L_1] \sim T[L_2]$, is that $\exists \gamma_T \in \mathbb{R}^1$ and $\Gamma_T \in \mathcal{G}$ such that $\forall L_o \in \mathcal{L}_o$, $\forall \omega \in \mathcal{W}$ one has

$$T[L_o * \omega] = T[L_o] * G_{0_k,\gamma_T,\Gamma_T}[\omega]$$

This actually reduces to $\exists \gamma_T \in \mathbb{R}^1$ such that $\forall L_o \in \mathcal{L}_o$, $\forall \beta \in \mathbb{R}^1_+$ one has

$$T[L_o * N(0_k, \beta I_k)] = T[L_o] * N(0_k, \gamma_T^2 \beta I_k)$$

Moreover, affine rotations distribute on $*$: $\forall L_1, L_2 \in \mathcal{L}$ one has

$$G_{v,\gamma,\Gamma}[L_1 * L_2] = G_{v,\gamma,\Gamma}[L_1] * G_{0_k,\gamma,\Gamma}[L_2]$$

It follows that *all affine rotations preserve equivalence*. For any invertible $G_{v,\gamma,\Gamma}[\cdot]$, i.e. whenever $\gamma \neq 0$ [11], one also has $L_o \in \mathcal{L}_o \Rightarrow G_{v,\gamma,\Gamma}[L_o] \in \mathcal{L}_o$. In fact, since the norm is

[10] The rotation that diagonalizes $\mathrm{Cov}(L)$, or equivalently the spectral basis, will not be unique if some of the eigenvalues have multiplicity larger than 1.

[11] Any $\Gamma \in \mathcal{G}$ is invertible.

unaffected by rotations, and $\Gamma'\Gamma = I_k$

$$\begin{aligned}\phi_{G_{v,\gamma,\Gamma}[L]}(u)\ e^{\frac{\beta}{2}\|u\|^2} &= e^{i(\gamma\Gamma u)'(\frac{1}{\gamma}\Gamma v)}\phi_L(\gamma\Gamma u) \\ &\quad e^{\frac{\beta}{2\gamma^2}\|\gamma\Gamma u\|^2} \\ &= e^{it'(\frac{1}{\gamma}\Gamma v)}\phi_L(t)\ e^{\frac{\beta}{2\gamma^2}\|t\|^2}\end{aligned}$$

setting $t = \gamma\Gamma u$. Hence, $\beta_o(G_{v,\gamma,\Gamma}[L]) = \gamma^2\beta_o(L)$; irreducibility of L implies irreducibility of $G_{v,\gamma,\Gamma}[L]$ (and vice-versa). Thus, *all invertible affine rotations are invariances* [12].

When restricting to a subspace S via marginalization

$$\mathcal{W}_S = \{N(0_k, \beta P_S), \beta \in \mathbb{R}^1_+\} = \{M_S[\omega], \omega \in \mathcal{W}\}$$

Moreover, also marginalizations distribute on $*$: $\forall\ L_1,\ L_2 \in \mathcal{L}$ one has

$$M_S[L_1 * L_2] = M_S[L_1] * M_S[L_2]$$

Due to Proposition 2.3, it follows that *all marginalizations preserve equivalence to any given $L_o \in \mathcal{L}_o$.* Hence, $\mathcal{M}(L_o)$ contains all and only the marginalizations preserving irreducibility of L_o itself. Equivalently, $\mathcal{M}(L_o)$ contains all and only the marginalizations such that $\forall L \sim L_o$, $\beta_{o,S}(M_S[L]) = \beta_o(L)$.

We also have that *all marginalizations are invariant on laws with bounded closed support.* In fact, if $\mathrm{Cs}(L_o)$ is bounded, under any marginalization $\mathrm{Cs}(M_S[L_o]) = P_S\mathrm{Cs}(L_o)$ is bounded, too, and therefore $M_S[L_o] \in \mathcal{L}_{o,S}$ by $OP(4)$ within S. Thus, if $\mathrm{Cs}(L_o)$ is bounded $\mathcal{M}(L_o)$ contains all marginalizations. Consequently (Proposition 2.5), $L_o \in \mathcal{K}$: *All laws with bounded closed support are kernels.*

Let us now see how the above setting could be put to work in practice. Suppose the law under investigation admits finite first and second order moments, and consider $G_L = G_{-\Gamma(L)E(L),1,\Gamma(L)}$, where $\Gamma(L) \in \mathcal{G}$ is a rotation diagonalizing $\mathrm{Cov}(L)$. Since this is an invariance, we can *transform to $G_L[L]$ before attempting the reduction.* The transformed law has

$$\mathrm{E}(G_L[L]) = 0_k \quad , \quad \mathrm{Cov}(G_L[L]) = \mathrm{Di}(\eta_i(L))$$

Correspondingly, the origin has

$$\mathrm{E}(RG_L[L]) = 0_k \quad , \quad \mathrm{Cov}(RG_L[L]) = \mathrm{Di}(\eta_i(L) - \beta_o(L))$$

Hence, indicating with $\{e_1, \ldots, e_k\}$ the standard orthonormal basis of $\mathbb{R}^k$, we can write

$$\begin{aligned}\mathbf{S}(RG_L[L]) &= \mathrm{Sp}(e_1, \ldots, e_{\nu(L)-1}) \\ &\oplus\ \mathrm{In}(\beta_o(L) < \eta_k(L))\ \mathrm{Sp}(e_{\nu(L)}, \ldots, e_k)\end{aligned}$$

This is a coordinate subspace: If $\beta_o(L) < \eta_k(L)$, it coincides with $\mathbb{R}^k$ (no dimension reduction is achieved). On the other hand, if $\beta_o(L) = \eta_k(L)$ the last $\nu(L)$ coordinates are eliminated.

Suppose one is further willing to assume that for the law under investigation $\mathrm{Cs}(R[L])$ is bounded. Clearly $\mathrm{Cs}(RG_L[L]) = \mathrm{Cs}(G_L R[L])$ is bounded, too. Thus, $RG_L[L]$ (as well

[12] Invertible affine rotations are also *the only affine invariances*: Let $A_{v,\gamma,A}$ indicate a generic invertible affine transformation, and suppose $L_2 = L_1 * \omega$. Distributivity still applies, so that $A_{v,\gamma,A}[L_2] = A_{v,\gamma,A}[L_1] * A_{0_k,\gamma,A}[\omega]$. But then, since $\mathcal{W}$ is closed *only* under centered affine rotations, $A_{v,\gamma,A}[L_1] \sim A_{v,\gamma,A}[L_2]$ requires $A \in \mathcal{G}$.

as $R[L]$) is a kernel, and we can *marginalize to the relevant coordinate subspace; that is, pass to* $M_{\mathbf{S}(RG_L[L])}G_L[L]$ *before attempting the reduction.*

In order to perform the above transformation and marginalization, we need to know $\mathrm{E}(L)$, $\Gamma(L)$, $\nu(L)$ and $\mathrm{In}(\beta_o(L) < \eta(L))$, or better, make inference on them based on data from L. Regarding $\mathrm{E}(L)$, and less trivially $\Gamma(L)$ and $\nu(L)$, several methods exist in the literature (see M.L. Eaton, D. Tyler, 1994, and E. Bura, 1997). Regarding $\mathrm{In}(\beta_o(L) < \eta(L))$, under the assumption that $\mathrm{Cs}(R[L])$ is bounded, we can again use the fact that $RG_L[L]$ is a kernel. As we remarked before, this gives $\beta_{o,S}(M_S G_L[L]) = \beta_o(G_L)$ for any choice of subspace S. Also, $\beta_o(G_L[L]) = \beta_o(L)$. Indicating with t the line given by $\mathrm{Sp}(e_k)$, we have then $\beta_{o,t}(M_t G_L[L]) = \beta_o(L)$, and $\mathrm{var}(M_t G_L[L]) = \eta_k(L)$ by construction. Down to one dimension, $\beta_{o,t}(M_t G_L[L]) = \mathrm{var}(M_t G_L[L])$ if and only if $M_t G_L[L]$ is normal. Thus, assessing whether $\beta_o(L) = \eta_k(L)$ (or conversely, strictly smaller), reduces to a normality check in one dimension (F. Chiaromonte, 1997).

In order to fully exploit the simplification potential of our framework, after transforming and marginalizing, one has to actually *reduce* $M_{\mathbf{S}(RG_L[L])}G_L[L]$. Notice this corresponds to a *de-convolution problem* [13] which has to be tackled in dimension possibly smaller than the original k. Notice also that if a drop in dimension indeed occured, the variability coefficient of the white noise to be de-convoluted from $M_{\mathbf{S}(RG_L[L])}G_L[L]$ is

$$\beta_{o,\mathbf{S}(RG_L[L])}(M_{\mathbf{S}(RG_L[L])}G_L[L]) = \beta_o(L) = \eta_k(L)$$

and can therefore be evaluated using an estimate of the smallest eigenvalue of $\mathrm{Cov}(L)$.

4 Final remarks

The convolution/white noises case constituted the obvious first step in the development of our framework: It represents a situation in which an independent and spherical normal error is additively superimposed to the "object of interest" [14]. The main aim of this paper is to show how this is just an instance of a very broad paradigm.

The general framework *does not rely on strong assumptions* on the nature of the law under investigation, and *allows us to produce a whole range of alternative definitions of structures, with reduction schemes identifying exhaustive simplified representations.*

The *definitions are rigorous* (as opposed to "intuitive"); correspondingly, the concepts of *simplification and exhaustiveness are unambiguous.* Moreover, each definition and reduction scheme *can be naturally interpreted* through the partial order, or binary operation and noise class, they arise from.

One could envision a situation in which *several definitions of structure are "fitted against" data from a given* L*, and the corresponding reductions are compared.* Some of the definitions might derive from (partial) orderings and/or superposition mechanisms and no-structure objects that are of particular relevance for the phenomenon at hand. Some could be easily mapped into incumbent or new theories on such phenomenon. Also, some of the definitions might induce larger dimension reductions than others.

Although the framework does not rely on strong assumptions, *the analysis is facilitated for laws whose reduction can be taken to be a kernel.* We have seen this in the convolution/white noises example, where boundedness of $\mathrm{Cs}(R[L])$ played a central role. It is important to remark that assumptions on $R[L]$ are inherently unverifiable, as data

[13] For an introduction to the issue, see A. Friedman, W. Littman, 1994 and G.M. Wing, 1991.

[14] Convolving two laws corresponds to adding two independent random vectors distributed according to them.

are drawn from L and not $R[L]$; likewise latent factors, $R[L]$ is *unobservable*. On the other hand, it is conceivable that for some definitions of structure $R[L] \in \mathcal{K}$ might be guaranteed through verifiable conditions on L. Moreover, even when the conditions must be posed on $R[L]$, they might still be quite reasonable on a large spectrum of practical applications (this is the case for boundedness of $\mathrm{Cs}(R[L])$, for example).

References

Buja, A., Cook, D., Swayne D.F. (1996), Interactive high-dimensional data visualization, *Journal of Computational and Graphical Statistics*, **v.5, n.1**, pp. 78–99.

Bura, E. (1997), Dimension reduction via parametric inverse regression, *L1 Statistical Procedures and Related Topics*, Y. Dodge (ed), Institute of Mathematical Statistics Lecture notes–monograph series, Hayward CA, U.S.A. 1997. pp. 215–228.

Chiaromonte, F. (1996), *A Reduction Paradigm for Multivariate Laws*, Ph.D. Dissertation, School of Statistics, University of Minnesota.

Chiaromonte, F. (1997), A reduction paradigm for multivariate laws, *L1 Statistical Procedures and Related Topics*, Y. Dodge (ed), Institute of Mathematical Statistics Lecture notes–monograph series, Hayward CA, U.S.A. 1997. pp. 229–240.

Chiaromonte, F., Cook R.D. (1997), On foundations of regression graphics, Technical Report n.616, School of Statistics, University of Minnesota.

Cook R.D. (1998), *Regression Graphics*, Wiley & Sons, New York (forthcoming).

Dawid, A.P. (1979), Conditional independence in statistical theory, *Journal of the Royal Statistical Society*, **ser.B, 41**, pp. 1–31.

Eaton, M.L. (1983), *Multivariate Statistics*, Wiley & Sons, New York.

Eaton, M.L., Tyler, D. (1994), The asymptotic distribution of singular values, with applications to canonical correlations and correspondence analysis, *Journal of Multivariate Analysis*, **v.50, n.2**.

Friedman, A., Littman, W. (1994), *Industrial Mathematics: a course in solving real world problems*, SIAM, Philadelphia.

Li, Ker Chau (1991), Sliced inverse regression for dimension reduction (with discussion), *Journal of the American Statistical Association*, **86**, pp. 316–342.

Whittaker, J. (1990), *Graphical Models in Applied Multivariate Statistics*, Wiley & Sons, New York.

Wing, G.M. (1991) *A Primer on Integral Equations of the First Kind: the Problem of Deconvolution and Unfolding*, SIAM, Port City Press.