L_1-*Statistical Procedures and Related Topics*
IMS Lecture Notes – Monograph Series (1997) Volume 31

A reduction paradigm for multivariate laws

Francesca Chiaromonte

International Institute for Applied Systems Analysis, Austria.

Abstract: A *reduction paradigm* is a theoretical framework which provides a definition of structure for multivariate laws, and allows to simplify their representation and statistical analysis. The main idea is to decompose a law as the superposition of a *structural term* and a *noise*, so that the latter can be neglected *without loss of information on the structure*. When the structural term is supported by a lower-dimensional affine subspace, an *exhaustive dimension reduction* is achieved. We describe the reduction paradigm that results from selecting white noises, and convolution as superposition mechanism.

Key words: Multivariate structure, multivariate dimension-reduction, multivariate graphics.

AMS subject classification: 62A99, 62H05.

1 Introduction

A k-variate law is a complex object whose structure embodies both marginal and joint features. All those features can be translated, to some extent, into geometric characterizations of an iid sample from the law, meant as a cloud of points in $\mathbb{R}^k$. Dimension does not affect the analysis of marginal features, but as k increases it becomes progressively harder to conceive and articulate the joint ones. For example, how does one conceive and articulate the interdependencies among, say, 10 or 100 coordinate components? One is often forced to neglect high-order interactions, and/or to assume hierarchies among them [1]. At the same time, for $k > 3$, the data cannot

[1] Conditional independence (see A.P. Dawid, 1979) provides a key to articulate interdependencies; a very interesting representation of them through *conditional independence*

be visualized as a whole; while graphical tools can still be used to investigate low-dimensional marginals, a *direct* graphical investigation of the joint features is impossible.

Producing inferences in high-dimensional settings can then become complicated and challenging. A large variety of inference methods is available once strong assumptions on the nature of the law are imposed; that is, once a model for the law is chosen (see, among others, M.L. Eaton, 1983, R.J. Muirhead, 1982, and G.A.F. Seber, 1984). But the intuition based on graphical preliminary exploration that should precede the utilization of model-based methods is impaired by the conceptual and practical difficulties mentioned above.

These considerations, among others, justify the quest for simplified representations of multivariate laws, especially ones allowing a *reduction in dimension*. Simplified representations are often developed targeting some (more or less restricted) features of interest. *Exhaustiveness* becomes then an issue; once a *target* has been chosen, the information concerning it ought to be preserved by simplification. More generally, it ought to be clear in what relation the proposed simplified representation is to the target. If exhaustiveness is not always guaranteed, it should be possible to state under what assumptions on the nature of the law it is, and/or to establish to what extent the target is preserved (with or without assumptions).

These issues are very relevant in practice; the last thirty years have witnessed the development of a large number of graphical exploration procedures for high-dimensional data sets. Think for example of Principal Component Analysis, Factor Analysis (see G.A.F. Seber, 1984, and references therein), Projection Pursuit (H.J. Friedman and J.W. Tuckey, 1974, H.J. Friedman, 1987, and D. Cook, A. Buja, J. Cabrera and C. Hurley, 1995), or Grand Tours (D. Asimov, 1985, and A. Buja and D. Asimov, 1986). The theoretical rationale underlying any of these procedures can be interpreted as a simplified representation of the multivariate law from which the data are drawn; targets range anywhere from "variability", to "linear interdependence structure" (correlation among the coordinate components), to "non-linear structure" (defined as departure from normality), to "structure" according to some other definition. Correspondingly, many of the critiques to these procedures can be interpreted in terms of choice of targets, and relations between simplified representations and targets. As we proceed, it will become clear that the simplified representation underlying Factor Analysis is the closest in spirit to the one we will propose. In fact, Factor Analysis differs from the other procedures mentioned above by

graphs is given by J. Whittaker, 1990.

its reference to a *latent factor* entirely embodying the correlation target.

Our focus will not be on techniques to make inference on simplified representations ("population" objects) based on data from a multivariate law, but on the theoretical premises for these techniques; that is, on how to define targets, and how to develop simplified representations guaranteed to embody them exhaustively.

In Sections 1 and 2, we introduce the concept of *reduction paradigm* and provide definitions and some key results. Section 3 concerns dimension reduction. We conclude with a brief summary and some remarks on inference in Section 4. More details can be found in F. Chiaromonte, 1996.

2 The reduction paradigm

Our analysis will be conducted at the level of laws on $\mathbb{R}^k$, and we will not distinguish among random vectors with the same distribution. The main idea behind a reduction paradigm is to decompose a law L on $\mathbb{R}^k$ into two terms, one of which does not contribute to the structure (the target) and can therefore be neglected. In other words, the aim is to represent a law as the superposition of a *structural term* and a *noise*, or *no-structure term*. Hence, the specification of a reduction paradigm relies upon

- a definition of absence of structure; that is, a choice of noises
- a choice of superposition mechanism

which, conversely, determine a definition of structure. We have selected *white noises* $N_k(0, \beta I_k)$, $\beta \in \mathbb{R}^1_+$, and *convolution*. Hence, we write

$$L = \Lambda_\beta(L) * N_k(0, \beta I_k) \tag{1}$$

or, in terms of characteristic functions

$$\phi_L(u) = \phi_{\Lambda_\beta(L)}(u) e^{-\frac{\beta}{2}\|u\|^2} ::, :: u \in \mathbb{R}^k \tag{2}$$

This is by no means the only possibility, but it is in line with much of the statistical tradition and thus constitutes a very natural first step. In fact, it expresses a situation in which an independent normal error is additively superimposed to the object of interest. One can envision reproducing the whole analysis we are about to develop with different noises and/or superposition mechanisms, though. As far as noises are concerned, one could take, for example, uniforms on hyper-spheres of radius $\rho \in \mathbb{R}^1_+$, or normals with independent components $N_k(0, \mathrm{Diag}(\sigma_i))$, $\sigma \in \mathbb{R}^k_+$. In the first case one maintains the weakly spherical nature of white noises and loses independence of the coordinate components, while in the second case one loses weak sphericity and maintains independence. Regarding superposition, one could explore, for example, multiplicative (instead of additive) schemes.

Before proceeding let us remark that the reduction paradigm we have selected, as well as any other conceivable one, while certainly constituting a *model for decomposing a law*, does not require strong assumptions on the nature of the law itself. A reduction paradigm can be applied without fixing at the outset a *model for the law*; that is, without assuming at the outset that the law belongs to a given (and possibly finally parameterized) class. Furthermore, our reduction paradigm corresponds to the inverse problem for heat-type diffusion of probability measures (for an easy introduction, see G.M. Wing, 1991, and A. Friedman and W. Littman, 1994). Paradigms resulting from a different choice of noises would correspond to inverse problems for processes with different kernels.

Indexing the structural term by β serves to stress the fact that the decomposition in (1) and (2) is not unique, unless it holds only with $\beta = 0$, and therefore $\Lambda_\beta(L) = L$ itself. The set

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

expresses the range of possible decompositions. It is always non-empty, as it must contain 0, and is easily shown to be $\mathcal{B}(L) = [0, \beta_o(L)]$, where $\beta_o(L) = \sup \mathcal{B}(L) = \max \mathcal{B}(L)$. We call the corresponding structural terms

$$\Lambda_\beta(L) \leftrightarrow \phi_L(\cdot)e^{\frac{\beta}{2}\|(\cdot)\|^2}, \ \beta \in \mathcal{B}(L)$$

sources, $\beta_o(L)$ *reduction coefficient*, and $\Lambda_o(L) \leftrightarrow \phi_L(\cdot)e^{\frac{\beta_o(L)}{2}\|(\cdot)\|^2}$ *primary source* of L. Notice that reduction coefficient and primary source are unique by construction. If $\beta_o(L) = 0$, so that the only (and thus primary) source of L is L itself, we say that the law is *irreducible*. We call it *reducible* otherwise.

All sources share the structure of L, and can be equivalently taken as exhaustive "simplified" representations of the law. The primary source is the one in which no error is superimposed to the structure; that is, the one in which we have pushed simplification as far as possible. Hence, we will select $\Lambda_o(L)$ as simplified representation of L, and write

$$L = \Lambda_o(L) * N_k(0, \beta_o(L) I_k)$$

We can fix ideas using the normal case as an example. Here and in the following, $P_{(\cdot)}$ indicates the orthogonal projection operator onto the argument subspace [2], with respect to the standard inner product on $\mathbb{R}^k$. Let

$$L = N_k\left(\mu, \sum_{j=1}^{p} \eta_j P_{V_j} + \eta P_V\right)$$

[2]The reference, throughout our discussion, is to *linear* subspaces, and affine subspaces obtained by translating them.

where $\sum_{j=1}^{p} \eta_j P_{V_j} + \eta P_V$ is the spectral decomposition of the covariance with (distinct) eigenvalues $\eta_1, \ldots, \eta_p, \eta$ in decreasing order, and corresponding eigenspaces $V_1, \ldots, V_p, V$. It is easy to show that

$$\phi_L(u) e^{\frac{\beta}{2}\|u\|^2} = \exp\left\{ iu'\mu - \frac{1}{2}u' \left(\sum_{j=1}^{p} (\eta_j - \beta) P_{V_j} + (\eta - \beta) P_V \right) u \right\}$$

is a characteristic function if and only if $\sum_{j=1}^{p} (\eta_j - \beta) P_{V_j} + (\eta - \beta) P_V$ is non-negative definite; that is, if and only if $\beta \leq \eta$. Hence, $\beta_o(L) = \eta$ and correspondingly

$$\Lambda_o(L) = N_k \left(\mu, \sum_{j=1}^{p} (\eta_j - \eta) P_{V_j} \right)$$

It is then clear that a normal is irreducible if and only if the smallest eigenvalue of its covariance is 0; the irreducible k-variate normals are all and only the ones supported by lower dimensional affine subspaces, and they constitute the primary sources of non-singular normals.

Primary sources are irreducible by construction. The class of all irreducible laws on $\mathbb{R}^k$ represents the repertoire of possible structures. The following proposition provides a *sufficient condition for irreducibility*, thereby characterizing part of such repertoire.

Proposition 1 *If there exists a measurable set* $B \subseteq \mathbb{R}^k$ *such that* $Leb(B) > 0$, *but* $L(B) = 0$, *then* L *is irreducible.*

Proof: Suppose $\beta_o(L) > 0$. Then, for any choice of $v \in \mathbb{R}^k$, $N_k(v, \beta_o(L) I_k)$ is mutually absolutely continuous with respect to *Leb*. So $Leb(B) > 0$ implies

$$N_k(v, \beta_o(L) I_k)(B) > 0, \ \forall v \in \mathbb{R}^k$$

and thus

$$L(B) = \int_{\mathbb{R}^k} N_k(v, \beta_o(L) I_k)(B) \ \Lambda_o(L)(dv) > 0$$

contradicting our assumption. We can conclude that $\beta_o(L) = 0$, and therefore that L is irreducible. □

Since we have selected white noises as no-structure terms, reducible laws must be mutually absolutely continuous with respect to the Lebesgue measure, because they "contain" a term that is. As a consequence, all laws having "thick" holes with respect to the Lebesgue measure are irreducible in $\mathbb{R}^k$. In particular, laws whose affine support $\mathrm{As}(L)$ has dimension $< k$ are irreducible in $\mathbb{R}^k$; we saw an instance of this with irreducible normals. So are laws whose closed support $\mathrm{Cs}(L)$ is bounded, regardless of whether

the latter is full-dimensional or embedded in a subspace or affine subspace of dimension $< k$.

Notice that existence of an everywhere positive density is *not* enough to guarantee reducibility; again because of our choice of no-structure terms, reducible laws' densities must have "thick enough" tails. It is easy to show that a law with an everywhere positive density whose tails vanish too fast, at least along some directions, will still be irreducible (see F. Chiaromonte, 1996).

3 Some affine actions, and marginalizations

We will now explore the effects on reduction of some affine actions and of marginalizations.

Proposition 2 *Let $\mathcal{T}_{v,r,R}[L]$ be the law of $rRX - v$, where $X \in \mathbb{R}^k$ is any random vector distributed according to L, $v \in \mathbb{R}^k$, $r \in \mathbb{R}^1$, and R is a rotation of $\mathbb{R}^k$. Then $\beta_o(\mathcal{T}_{v,r,R}[L]) = \beta_o(L)$ and $\Lambda_o(\mathcal{T}_{v,r,R}[L]) = \mathcal{T}_{v,r,R}[\Lambda_o(L)]$.*

Proof: For $r = 0$, the transformation yields a point-mass at $-v$, and the statement is trivially true. Otherwise, using characteristic functions, one has

$$\begin{aligned}
\phi_{\mathcal{T}_{v,r,R}[L]}(u) &= e^{-\mathrm{i}u'v}\phi_L(rR'u) \\
&= e^{-\mathrm{i}u'v}\phi_{\Lambda_o(L)}(rR'u)e^{-\frac{\beta_o(L)}{2}\|rR'u\|^2} \\
&= \phi_{\mathcal{T}_{v,r,R}[\Lambda_o(L)]}(u)e^{-\frac{r^2\beta_o(L)}{2}\|u\|^2}
\end{aligned}$$

so $\beta_o(\mathcal{T}_{v,r,R}[L]) \geq r^2\beta_o(L)$, and $\mathcal{T}_{v,r,R}[\Lambda_o(L)]$ is a source of $\mathcal{T}_{v,r,R}[L]$. But for $r \neq 0$ our transformation is invertible: $\mathcal{T}^{-1}_{v,r,R}[\cdot] = \mathcal{T}_{-v,1/r,R'}[\cdot]$. Hence

$$\begin{aligned}
\phi_L(u) &= \phi_{\mathcal{T}_{-v,1/r,R'}\mathcal{T}_{v,r,R}[L]}(u) \\
&= \phi_{\mathcal{T}_{-v,1/r,R'}[\Lambda_o(\mathcal{T}_{v,r,R}[L])]}(u)e^{-\frac{(1/r)^2\beta_o(\mathcal{T}_{v,r,R}[L])}{2}\|u\|^2}
\end{aligned}$$

and $\beta_o(L) \geq (1/r)^2\beta_o(\mathcal{T}_{v,r,R}[L])$. We can conclude that $\beta_o\ (\mathcal{T}_{v,r,R}[L]) = r^2\ \beta_o(L)$, and therefore that $\mathcal{T}_{v,r,R}[\Lambda_o(L)]$ is indeed the primary source of $\mathcal{T}_{v,r,R}[L]$. □

The reduction coefficient is not affected by rotations and translations, and is multiplied by the square of a rescaling factor. Thus, rescalings, rotations and translations of L result into corresponding rescalings, rotations and translations of the primary source. In the following, we will use

interchangeably the terms *marginalization* and *projection.* Besides the intuitive correspondence, "invariance" under rotations makes this rigorous; the choice of orthonormal basis does not matter.

In our discussion so far, we have considered the reduction of a law L on $\mathbb{R}^k$ *in* $\mathbb{R}^k$. The reference to the space is important; laws on $\mathbb{R}^k$ that are entirely concentrated on some subspace can also be meant as laws on such subspace, and reducing them *within the subspace* can produce a different set of sources, a different reduction coefficient and a different primary source. Laws that are entirely concentrated on a subspace of dimension $< k$ are irreducible in $\mathbb{R}^k$, but they might still be reducible within the subspace.

The noises within a given subspace $S \subseteq \mathbb{R}^k$ are represented by $N_k(0, \beta P_S)$, $\beta \in \mathbb{R}_+^1$. Notation-wise, when considering the reduction of a law $\tilde{L}$ (entirely concentrated on S) within S, we will write $\beta_o(\tilde{L}, S)$, $\Lambda_o(\tilde{L}, S)$, etc.

Proposition 3 *Let $\mathcal{M}_S[L]$ be the law of $P_S X$, where $X \in \mathbb{R}^k$ is any random vector distributed according to L, and $S \subseteq \mathbb{R}^k$ is a non-degenerate subspace. Then $\beta_o(\mathcal{M}_S[L], S) \geq \beta_o(L)$ and*

$$\Lambda_o(\mathcal{M}_S[L], S) * N_k(0, \alpha P_S) = \mathcal{M}_S[\Lambda_o(L)]$$

where $\alpha = \beta_o(\mathcal{M}_S[L], S) - \beta_o(L)$. In particular, if $Cs(\Lambda_o(L))$ is bounded, $\beta_o(\mathcal{M}_S[L], S) = \beta_o(L)$ and $\Lambda_o(\mathcal{M}_S[L], S) = \mathcal{M}_S[\Lambda_o(L)]$.

Proof: Using characteristic functions, one has

$$\begin{aligned} \phi_{\mathcal{M}_S[L]}(u) &= \phi_L(P_S u) \\ &= \phi_{\Lambda_o(L)}(P_S u) e^{-\frac{\beta_o(L)}{2}\|P_S u\|^2} \\ &= \phi_{\mathcal{M}_S[\Lambda_o(L)]}(u) e^{-\frac{\beta_o(L)}{2}\|P_S u\|^2} \end{aligned}$$

so $\beta_o(\mathcal{M}_S[L], S) \geq \beta_o(L)$, and $\mathcal{M}_S[\Lambda_o(L)]$ is a source of $\mathcal{M}_S[L]$ within S. Equating the right hand side above with the right hand side of

$$\phi_{\mathcal{M}_S[L]}(u) = \phi_{\Lambda_o(\mathcal{M}_S[L],S)}(u) e^{-\frac{\beta_o(\mathcal{M}_S[L],S)}{2}\|P_S u\|^2}$$

we obtain

$$\phi_{\Lambda_o(\mathcal{M}_S[L],S)}(u) e^{-\frac{\alpha}{2}\|P_S u\|^2} = \phi_{\mathcal{M}_S[\Lambda_o(L)]}(u)$$

where $\alpha = \beta_o(\mathcal{M}_S[L], S) - \beta_o(L)$; that is

$$\Lambda_o(\mathcal{M}_S[L], S) * N_k(0, \alpha P_S) = \mathcal{M}_S[\Lambda_o(L)]$$

Now, assume that $\mathrm{Cs}(\Lambda_o(L))$ is bounded. We need to show that this implies $\alpha = 0$. Suppose $\alpha > 0$. Then, because of the normal term

$$\mathrm{Cs}(\mathcal{M}_S[\Lambda_o(L)]) = \mathrm{Cs}(\Lambda_o(\mathcal{M}_S[L], S) * N_k(0, \alpha P_S)) = S$$

But if $\mathrm{Cs}(\mathcal{M}_S[\Lambda_o(L)])$ is unbounded $\mathrm{Cs}(\Lambda_o(L))$ must be unbounded, too, contradicting our assumption. We can conclude that $\beta_o(\mathcal{M}_S[L], S) = \beta_o(L)$, and therefore that $\Lambda_o(\mathcal{M}_S[L], S) = \mathcal{M}_S[\Lambda_o(L)]$. □

The reduction coefficient (within S) of the marginal of L, must be greater than or equal to $\beta_o(L)$. Correspondingly, the marginal of $\Lambda_o(L)$ is a source (within S) of the marginal of L, even though not necessarily the primary one. Under the assumption that $\mathrm{Cs}(\Lambda_o(L))$ is bounded, the reduction coefficients coincide. Thus, the marginal of $\Lambda_o(L)$ is indeed the primary source (within S) of the marginal of L. In other words, under the boundedness assumption the reduction coefficient is not affected by marginalizations (projections), and therefore marginalizations of L result into corresponding marginalizations of the primary source.

4 The structural subspace, and exhaustive dimension reduction

The affine support of $\Lambda_o(L)$ represents the smallest affine subspace supporting the structure of L, as defined by our reduction paradigm. We call the subspace underlying it the *structural subspace* of the law $S_o(L) = \mathrm{As}(\mathcal{T}_v[\Lambda_o(L)])$, where v is any element of $\mathrm{Cs}(\Lambda_o(L))$, and $\mathcal{T}_v$ stands for $\mathcal{T}_{v,1,I_k}$. Correspondingly, we call $d_o(L) = \dim(S_o(L))$ the *structural dimension.* Whenever $d_o(L) < k$, our (exhaustive) simplified representation of L implies a drop in dimension.

This allows us to define an *exhaustive dimension reduction.* Let us see how. Suppose we know $v \in \mathrm{Cs}(\Lambda_o(L))$. Then, the exercise of identifying $\Lambda_o(L)$ is equivalent to that of identifying $\Lambda_o(\mathcal{T}_v[L])$. In fact, by Proposition 2

$$\Lambda_o(L) = \mathcal{T}_{-v}\mathcal{T}_v[\Lambda_o(L)] = \mathcal{T}_{-v}[\Lambda_o(\mathcal{T}_v[L])]$$

Now, suppose $S_o(L)$ is known, too. Then, we can marginalize $\mathcal{T}_v[L]$ to $S_o(L)$ preserving all the information relative to the structure, as defined by our reduction paradigm. In fact, again by Proposition 2

$$S_o(L) = \mathrm{As}(\mathcal{T}_v[\Lambda_o(L)]) = \mathrm{As}(\Lambda_o(\mathcal{T}_v[L]))$$

so that indeed $\Lambda_o(\mathcal{T}_v[L])$ is supported by the structural subspace [3], and

$$\Lambda_o(\mathcal{T}_v[L]) = \mathcal{M}_{S_o(L)}[\Lambda_o(\mathcal{T}_v[L])]$$

[3]Notice that, by Proposition 2, the structural subspace is invariant under translations of L : $S_o(\mathcal{T}_v[L]) = S_o(L)$. When translating by an element of $\mathrm{Cs}(\Lambda_o(L))$ we obtain a law which is actually supported by the subspace itself, instead of an affine subspace parallel to it.

But then, by Proposition 3

$$\Lambda_o(\mathcal{T}_v[L]) = \Lambda_o(\mathcal{M}_{S_o(L)}\mathcal{T}_v[L], S_o(L)) * N_k(0, \alpha P_{S_o(L)})$$

$\Lambda_o(\mathcal{T}_v[L])$ is a source for $\mathcal{M}_{S_o(L)}\mathcal{T}_v[L]$ within $S_o(L)$. Furthermore, if we can assume $\mathrm{Cs}(\Lambda_o(L))$, and therefore of $\mathrm{Cs}(\Lambda_o(\mathcal{T}_v[L]))$, to be bounded

$$\Lambda_o(\mathcal{T}_v[L]) = \Lambda_o(\mathcal{M}_{S_o(L)}\mathcal{T}_v[L], S_o(L))$$

that is, $\Lambda_o(\mathcal{T}_v[L])$ is the primary source of $\mathcal{M}_{S_o(L)}\mathcal{T}_v[L]$ within $S_o(L)$. This gives an even stronger meaning to the exhaustiveness of our marginalization; not only no structural information is lost marginalizing $\mathcal{T}_v[L]$, but the exercise of identifying $\Lambda_o(\mathcal{T}_v[L])$ (to be performed in k dimensions) would actually correspond to that of identifying $\Lambda_o(\mathcal{M}_{S_o(L)}\mathcal{T}_v[L], S_o(L))$ (to be performed in –possibly– smaller dimension).

The question becomes then how to identify translation term and structural subspace. Clearly, existence of finite moments of a certain order for L implies that of the corresponding moments for $\Lambda_o(L)$. If L admits finite first order moments $\mathrm{E}(\Lambda_o(L)) = \mathrm{E}(L)$, and one can take as translation term $v = \mathrm{E}(L) \in \mathrm{Cs}(\Lambda_o(L))$. Furthermore, if L admits finite second order moments, structural subspace and structural dimension can be related to the spectral decomposition of the covariance. We will need the following

Lemma 1 *If L admits finite second order moments, then $As(\mathcal{T}_{E(L)}[L]) = Span(Cov(L))$.*

Proof: Consider the orthogonal complement of $\mathrm{Span}(\mathrm{Cov}(L))$ with respect to the standard inner product, and $\mathcal{M}_{\mathrm{Span}(\mathrm{Cov}(L))^\perp}\mathcal{T}_{\mathrm{E}(L)}[L]$. Easy calculations give

$$\mathrm{E}(\mathcal{M}_{\mathrm{Span}(\mathrm{Cov}(L))^\perp}\mathcal{T}_{\mathrm{E}(L)}[L]) = 0,\ \mathrm{Cov}(\mathcal{M}_{\mathrm{Span}(\mathrm{Cov}(L))^\perp}\mathcal{T}_{\mathrm{E}(L)}[L]) = 0$$

Thus, $\mathcal{M}_{\mathrm{Span}(\mathrm{Cov}(L))^\perp}\mathcal{T}_{\mathrm{E}(L)}[L]$ is a point-mass at 0, which implies $\mathrm{Span}(\mathrm{Cov}(L)) \supseteq \mathrm{As}(\mathcal{T}_{\mathrm{E}(L)}[L])$. On the other hand, using the definition of covariance we have

$$\mathrm{Cov}(L)z = \mathrm{Cov}(\mathcal{T}_{\mathrm{E}(L)}[L])z = \int_{\mathrm{Cs}(\mathcal{T}_{\mathrm{E}(L)}[L])} vv'z\ \mathcal{T}_{\mathrm{E}(L)}[L](dv)$$

which implies $\mathrm{Span}(\mathrm{Cov}(L)) \subseteq \mathrm{As}(\mathcal{T}_{\mathrm{E}(L)}[L])$. The statement follows. □

Now, denoting by $\mathrm{Ind}(\cdot)$ the indicator function of the argument condition, we have

Proposition 4 *Suppose L admits finite second order moments. Let $\eta(L)$ be the smallest eigenvalue of $Cov(L)$, and $V(L)$ the corresponding eigenspace. Then $\beta_o(L) \leq \eta(L)$ and*

$$S_o(L) = V(L)^{\perp} \oplus Ind(\eta(L) - \beta_o(L) > 0)V(L)$$

with $d_o(L) = k - Ind(\eta(L) - \beta_o(L) = 0)dim(V(L))$.

Proof: Writing $\mathrm{Cov}(L) = \sum_{j=1}^{p} \eta_j(L)P_{V_j(L)} + \eta(L)P_{V(L)}$ one has

$$\begin{aligned} \mathrm{Cov}(\Lambda_o(L)) &= \mathrm{Cov}(L) - \beta_o(L)I_k \\ &= \sum_{j=1}^{p}(\eta_j(L) - \beta_o(L))P_{V_j(L)} + (\eta(L) - \beta_o(L))P_{V(L)} \end{aligned}$$

But then $\beta_o(L) \leq \eta(L)$ is implied by non-negative definiteness of $\mathrm{Cov}(\Lambda_o(L))$, and the expression for the structural subspace follows from Lemma 1 applied to $\Lambda_o(L)$. A drop in dimension occurs if and only if $\beta_o(L) = \eta(L)$, and when it occurs $d_o(L) = k - \dim(V(L))$, where $\dim(V(L))$ represents the multiplicity of $\eta(L)$. □

Given the spectral decomposition of $\mathrm{Cov}(L)$, the above proposition provides an upper bound for the reduction coefficient and a lower bound for the structural subspace; namely, the smallest eigenvalue of $\mathrm{Cov}(L)$ and the orthogonal complement of its eigenspace. The spectral decomposition of $\mathrm{Cov}(L)$ is not enough to identify the structural subspace, though; we still need to know whether the reduction coefficient is strictly smaller than, or equal to, the smallest eigenvalue.

Remember that for a normal law $\beta_o(L) = \eta$. Thus, under normality the drop in dimension always occurs, and one has $S_o(L) = V^{\perp}$ with $d_o(L) = k - \dim(V) \leq k-1$. It is important to remark that coincidence of $\beta_o(L)$ with $\eta(L)$ (and therefore the drop in dimension) is not guaranteed in general. Identifying the reduction coefficient with the smallest eigenvalue of the covariance can actually be very misleading. Take for example a "noisy" uniform on a hyper-cube $L = Un([-\theta,\theta]^k) * N_k(0,\tau I_k)$, $\theta,\tau \in \mathbb{R}_+^1 \setminus \{0\}$. For such a law one has $\beta_o(L) = \tau < \frac{\theta^2}{2} + \tau = \eta(L)$ and $S_o(L) = \mathbb{R}^k \supset \{0\} = V(L)^{\perp}$, as the multiplicity of $\frac{\theta^2}{2} + \tau$ is k.

5 A brief summary with some remarks on inference

The ultimate aim within the framework defined by a reduction paradigm is that of making inference about the (unobservable) $\Lambda_o(L)$, which constitutes

our simplified and yet exhaustive representation of the original law. As we have seen, if we can assume that L is normal, $\beta_o(L) = \eta$ and

$$\Lambda_o(L) = N_k\left(\mu, \sum_{j=1}^{p}(\eta_j - \eta)P_{V_j}\right)$$

which is entirely identified through the mean vector and the spectral decomposition of the covariance. Hence, if the data are consistent with normality, we could estimate the primary source based on estimates of those. Also, if the data can be transformed to approximate normality, the primary source could be estimated on the transformed scale. What can we do when the data contradicts normality on the original scale, and fails to approximate it also after applying normalizing transformations?

An intermediate aim is constituted by estimating a $v \in \mathrm{Cs}(\Lambda_o(L))$ and $S_o(L)$. Besides the intrinsic interest, if indeed our simplified representation implied a drop in dimension, having such estimates would allow us to perform an exhaustive dimension reduction.

Given the results in the previous sections, we are clearly at an advantage if we are willing to assume boundedness of $\mathrm{Cs}(\Lambda_o(L))$. Since the latter implies existence and finiteness for all the moments of L, we would have $\mathrm{E}(L) \in \mathrm{Cs}(\Lambda_o(L))$ and (Proposition 4)

$$S_o(L) = V(L)^{\perp} \oplus \mathrm{Ind}(\eta(L) - \beta_o(L) > 0)V(L)$$

Furthermore, we could restrict inference on the reduction coefficient to any arbitrarily small non-degenerate subspace. In fact, by Proposition 3 $\beta_o(L) = \beta_o(\mathcal{M}_t[L], t)$, where t is any line in $\mathbb{R}^k$. Thus, we could take $\hat{\mathrm{E}}(L)$ as translation term, and produce an estimate of the structural subspace based on $\hat{\eta}(L)$, $\hat{V}(L)$ and $\hat{\beta}_o(\mathcal{M}_t[L], t)$. Methods to estimate $\mathrm{E}(L)$, and, less trivially, $\eta(L)$ and $V(L)$, exist in the literature and are not affected by how large k is (see M.L. Eaton and D. Tyler, 1994, and E. Bura, 1996).

As a matter of fact, in order to produce an estimate of the structural subspace we would only have to assess, selecting for example $t \subseteq \hat{V}$, whether $\beta_o(\mathcal{M}_t[L], t)$ is strictly smaller than, or coincides with $\mathrm{var}(\mathcal{M}_t[L]) = \eta(L)$. This, in turn, is equivalent to assessing whether $\mathcal{M}_t[L]$ is a 1-dimensional normal.

Under the assumption that $\mathrm{Cs}(\Lambda_o(L))$ is bounded, we also have that

$$\Lambda_o(L) = \mathcal{T}_{-v}[\Lambda_o(\mathcal{M}_{S_o(L)}\mathcal{T}_v[L], S_o(L))]$$

Hence, we could center the data cloud translating it by $\hat{\mathrm{E}}(L)$, and restrict any further analysis to the projection of the centered cloud onto $\hat{S}_o(L)$;

all the structural features (except for location, which is captured by $\hat{E}(L)$) would be preserved. If indeed $\hat{d}_o(L) = \dim(\hat{S}_o(L)) < k$, we would have achieved an exhaustive dimension reduction.

References

[1] Asimov D. (1985). The grand tour: a tool for viewing multidimensional data. *SIAM J. Scient. Statist. Comput.* **6**, 128–143.

[2] Buja A., and Asimov D. (1986). Grand tour methods: an outline. *Computing Science and Statistics* **17**, 63–67.

[3] Bura E. (1996). Dimension reduction via inverse regression. Ph.D. Dissertation, School of Statistics, University of Minnesota.

[4] Chiaromonte F. (1996). A reduction paradigm for multivariate laws. Ph.D. Dissertation, School of Statistics, University of Minnesota.

[5] Cook D., Buja A., Cabrera J., and Hurley C. (1995). Grand tour and projection pursuit. *J. Comput. Graph. Statist.* **4** 155–172.

[6] Dawid A.P. (1979). Conditional independence in statistical theory. *J. R. Statist. Soc.*B **41**, 1–31.

[7] Eaton M.L. (1983). *Multivariate Statistics.* New York: Wiley.

[8] Eaton M.L., and Tyler D. (1994). The asymptotic distribution of singular values, with applications to canonical correlations and correspondence analysis. *J. Mult. Anal.* **50**.

[9] Friedman A., and Littman W. (1994). *Industrial Mathematics: a Course in Solving Real World Problems.* Philadelphia: SIAM.

[10] Friedman H.J. (1987). Exploratory projection pursuit. *J. Am. Statist. Assoc.* **82**, 249–266.

[11] Friedman H.J., and Tuckey J.W. (1974). A projection pursuit algorithm for exploratory data analysis. *IEEE Transactions on Computers*C **23**, 881–889.

[12] Muirhead R.J. (1982). *Aspects of Multivariate Statistical Theory.* New York: Wiley.

[13] Posse C. (1995). Tools for two-dimensional exploratory projection pursuit. *J. Comput. Graph. Statist.***4**, 83–100.

[14] Seber G.A.F. (1984). *Multivariate Observations.* New York: Wiley.

[15] Whittaker J. (1990). *Graphical Models in Applied Multivariate Statistics.* New York: Wiley.

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