# Independent Subspaces

**Lei Xu**

*Chinese University of Hong Kong, Hong Kong, & Peking University, Beijing, China*

## INTRODUCTION

Several unsupervised learning topics have been extensively studied with wide applications for decades in the literatures of statistics, signal processing, and machine learning. The topics are mutually related and certain connections have been discussed partly, but still in need of a systematical overview. The article provides a unified perspective via a general framework of independent subspaces, with different topics featured by differences in choosing and combining three ingredients. Moreover, an overview is made via three streams of studies. One consists of those on the widely studied principal component analysis (PCA) and factor analysis (FA), featured by the second order independence. The second consists of studies on a higher order independence featured independent component analysis (ICA), binary FA, and nonGaussian FA. The third is called mixture based learning that combines individual jobs to fulfill a complicated task. Extensive literatures make it impossible to provide a complete review. Instead, we aim at sketching a roadmap for each stream with attentions on those topics missing in the existing surveys and textbooks, and limited to the authors' knowledge.

## A GENERAL FRAMEWORK OF INDEPENDENT SUBSPACES

A number of unsupervised learning topics are featured by its handling on a fundamental task. As shown in Fig.1(b), every sample $x$ is projected into $\hat{x}$ on a manifold and the error $e = x - \hat{x}$ of using $\hat{x}$ to represent $x$ is minimized collectively on a set of samples. One widely studied situation is that a manifold is a subspace represented by linear coordinates, e.g., spanned by three linear independent basis vectors $a_1, a_2, a_3$ as shown in Fig.1(a). So, $\hat{x}$ can be represented by its projection $y^{(j)}$ on each basis vector, i.e.,

$$\hat{x} = \sum_{j}^{3} y^{(1)} a_j$$

or

$$x = \hat{x} + e = Ay + e, \quad [y = \quad y^{(1)}, y^{(2)}, y^{(3)}]^T. \tag{1}$$

Typically, the error $e = x - \hat{x}$ is measured by the square norm, which is minimized when $e$ is orthogonal to $\hat{x}$. Collectively, the minimization of the average error $\|e\|^2$ on a set of samples or its expectation $E\|e\|^2$ is featured by those natures given at the bottom of Fig.1(a).

Generally, the task consists of three ingredients, as shown in Fig.2. First, how the error $e = x - \hat{x}$ is measured. Different measures define different projections. The square norm $d = \|e\|^2$ applies to a homogeneous medium between $x$ and $\hat{x}$. Other measures are needed for inhomogeneous mediums. In Fig.1(c), a non-orthogonal but still linear projection is considered via $d = \|e\|_B^2 = e^T \Sigma_e^{-1} e$ with $\Sigma_e^{-1} = B^T B$, as if $e$ is first mapped to a homogeneous medium by a linear mapping $e$ and then measured by the square norm. Shown at the bottom of Fig.1(c) are the natures of this $Min\|e\|_B^2$. Being considerably different from those of $Min\|e\|^2$, more assumptions have to be imposed externally.

The second ingredient is a coordinate system, via either linear vectors in Fig.1(a)&(c) or a set of curves on a nonlinear manifold in Fig.1(b). Moreover, there is the third ingredient that imposes certain structure to further constrict how $y$ is distributed within the coordinates, e.g., by the nature d).

The differences in choosing and combining the three ingredients lead to different approaches. We use the name "independent subspaces" to denote those structures with the components of $y$ being mutually independent, and get a general framework for accommodating several unsupervised learning topics.

Subsequently, we summarize them via three streams of studies by considering

- $d = \|e\|_B^2 = e^T \Sigma_e^{-1} e$ and two special cases,
- three types of independence structure, and whether there is temporal structure among samples,
- varying from one linear coordinate system to multiple linear coordinate systems at different locations, as shown in Fig.2.
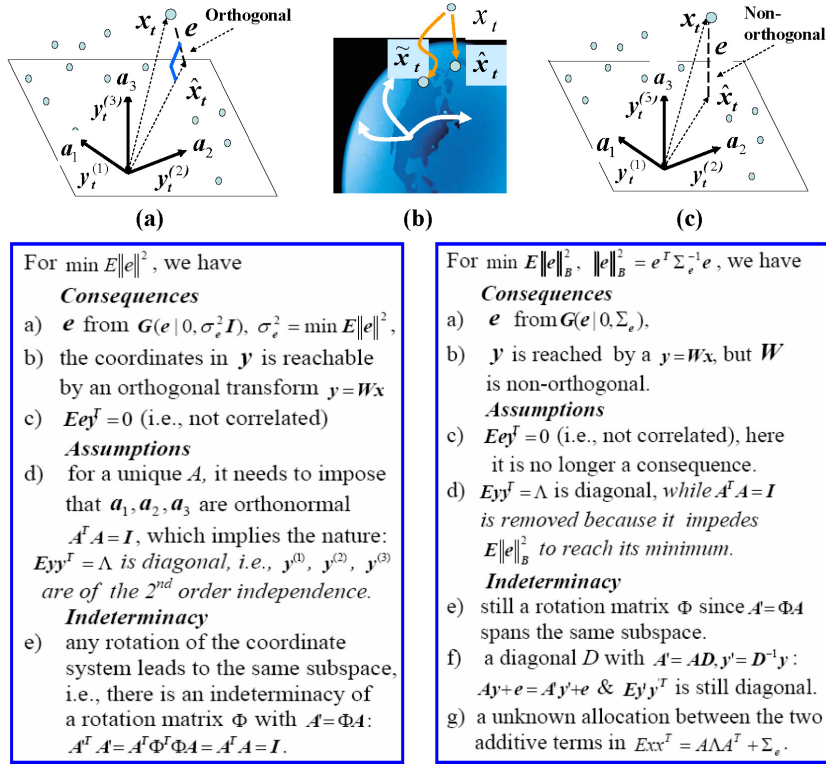
*Figure 1.*

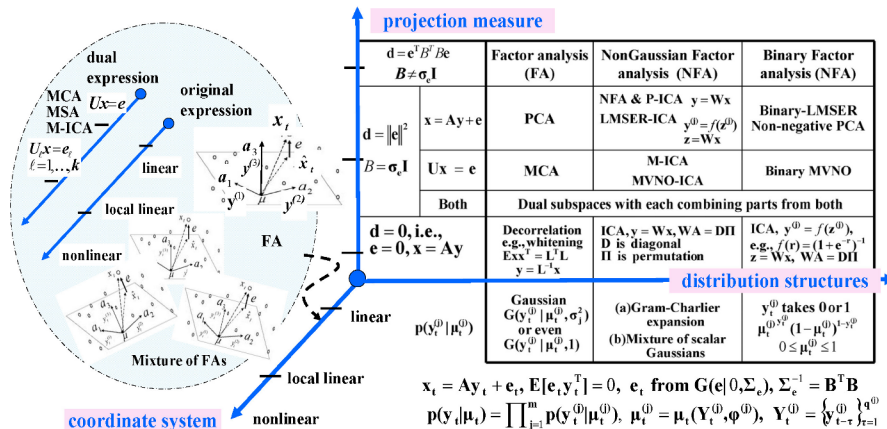**Fig.1  A General Framework of Independent Subspaces**

*Figure 2.*

**Fig.2  Three gradients and their typical choices**

## Related Content

B-School Selection by Fuzzy TOPSIS and AHP

Vivek Agrawal, Vikas Tripathiand Nitin Seth (2018). *Intelligent Systems: Concepts, Methodologies, Tools, and Applications  (pp. 929-955).*

www.irma-international.org/chapter/b-school-selection-by-fuzzy-topsis-and-ahp/205816

Modeling Gender Based Customer Preferences of Information Search Channels

Gaurav Khatwaniand Praveen Ranjan Srivastava (2018). *Intelligent Systems: Concepts, Methodologies, Tools, and Applications  (pp. 622-638).*

www.irma-international.org/chapter/modeling-gender-based-customer-preferences-of-information-search-channels/205801

Hyper-Personalization: An AI-Enabled Personalization for Customer-Centric Marketing

Darshana Desai (2022). *Adoption and Implementation of AI in Customer Relationship Management (pp. 40-53).*

www.irma-international.org/chapter/hyper-personalization/289446

Educational Technology and Libraries Supporting Online Learning

 Kumar N, S. Antoniraj, S. Jayanthi, S. Mirdula, Saravanakumar Selvaraj, N. Rajkumarand K. R. Senthilkumar (2024). *AI-Assisted Library Reconstruction (pp. 209-237).*

www.irma-international.org/chapter/educational-technology-and-libraries-supporting-online-learning/343588

EnOntoModel: A Semantically-Enriched Model for Ontologies

Nwe Ni Tunand Satoshi Tojo (2010). *Methodological Advancements in Intelligent Information Technologies: Evolutionary Trends  (pp. 128-158).*

www.irma-international.org/chapter/enontomodel-semantically-enriched-model-ontologies/38525