

# Like Mother, like daughter?

## A dyadic sequence analysis approach to uncover patterns of mothers and daughters careers

**Authors:** Nicolas Robette, Printemps (UVSQ-CNRS, UMR 8085)  
Xavier Bry, I3M, Université Montpellier 2  
Éva Lelièvre, INED

### Abstract:

The analysis of the occupational trajectories of successive generations i.e. parents and their children, set aside the availability of suitable data, requires descriptive tools which allow to cross and match the information on the similarities/dissimilarities within a particular generation to those of the next and then gives a synthetic overview of the long term inter-generational patterns. This paper presents the findings of a study which develops a “dyadic sequence analysis” approach (DSA), unfolding in stages: starting with an optimal matching analysis (OMA) of the employment histories of French women born between 1930 and 1950 and that of their mothers, born between 1885 and 1935, we reduce the dissimilarity matrices by Multidimensional Scaling (MDS) and finally present the intra-family dyads of female careers formed by sequences of a mother's work history and that of her daughter using clustering based on structural links between the mothers' and daughters' careers (i.e. canonical PLS).

### Keywords:

Career, female employment, France, intergenerational transmission, optimal matching, sequence analysis

## 1. Introduction

Sequence analysis of work careers and social trajectories in general is now well established (Brzinsky-Fay and Kohler, 2010). Notable previous developments have tended to explore the possibility to take into account not only employment statuses but also integrate multivariable combinations of statuses pertaining to other dimensions of the life-courses such as housing and family, introducing multidimensional sequence analysis (Pollock, 2007). In this paper, we examine another aspect of occupational trajectories: that of social mobility from one generation to the next. While social mobility is more usually studied in terms of cross-sectional comparison of the position of the father at a given period compared to the position of the son, we here consider the work histories of both generations. More precisely, we here study women's involvement in paid employment for two successive generations marked, in France, by their massive entry into the labour force (Beller, 2009).

To do so, we will first outline the characteristic patterns of the work histories of women born between 1930 and 1950, through a typology obtained by sequence analysis and clustering. We will then compare those to the career profiles of their mothers and end by describing the intra-family dyads of female careers formed by mothers' work history and that of their daughters. For this we use an approach based on structural links between the mothers' and daughters' careers, combining Optimal Matching Analysis (OMA), Multidimensional Scaling (MDS), canonical Partial Least Square (PLS) and clustering, an approach we call "Dyadic Sequence Analysis" (DSA).

To examine the *long term* evolution we here take advantage of the 2001 INED<sup>1</sup> survey *Biographies et entourage* from which working careers can be tracked over three generations (the respondents, their parents and their children), and we proceed by stages to identify the inter-generational profiles of women's employment patterns. In this paper we precisely deal with the occupational trajectories of women respondents born between 1930 and 1950 in continuity with those of their mothers, born between 1885 and 1935.

---

<sup>1</sup> Institut National d'Etudes Démographiques

## 2. A first issue for inter-generational work histories: the data

INED's *Biographies et entourage* is a retrospective life-event history survey of 2830 residents of the Île-de-France<sup>2</sup> area aged between 50 and 70, and those of their contact circles. The sample interviewed was representative of the Île-de-France region population in 2000, the year the survey data were collected. The 'contact circle' (*entourage*) includes family members (lineal kin and relations by marriage) across four generations, plus all those with whom the respondent had cohabited and any others who, related or not, had played a major role in their life (Lelièvre and Vivier, 2001, GRAB, 2009). The *Biographies et entourage* questionnaire recorded all stages in the residential, occupational and family trajectories of the respondents and the members of their contact circle, year by year using an extended Life History Calendar. It is thus possible to reconstitute individual trajectories in detail and to consider the work histories of more than one generation, notably the female respondents and their mothers.

Gathering life course data retrospectively over a long period and by proxy (getting respondents to recount their parents' or children's event histories) required appropriate tools. The *Biographies et entourage* survey had the benefit of know-how accumulated from earlier life course studies (GRAB, 1999). The GRAB<sup>3</sup> studies allowed respondent to contextualise their narratives by correlating family, residential and occupational calendars by means of a grid (a record sheet called *Ageven*<sup>4</sup>). This data collecting method replaces the usual question-and-answer system and has proved well suited to recording a narrative in which events in different spheres are called to mind in a more natural way, by reference to each other.

We thus have the whole of a respondent's occupational history including periods of economic inactivity (only periods of more than a year were included). Each stage is characterised by the occupation declared by the respondent, their employment status, a description of the employer (public or private sector, economic branch, location, size) and the proportion of time spent at work. For the work histories of the respondent's parents, less precision was required but we were able to reconstruct their work positions and job types, as well as any career interruptions and their nature.

---

<sup>2</sup> The Île-de-France is constituted by the Paris metropolitan area plus the outer suburbs, a Region concentrating 19% of the French population in 2000.

<sup>3</sup> Groupe de Réflexion sur l'Approche Biographique

<sup>4</sup> *Ageven* for "*âge et évènement*" (i.e. age and event) introduced in 1987 by Antoine *et al.*.

### 3. Family and occupational trajectories of women born between 1930 and 1950 in France

The individual trajectories of the women respondents are spread over a 70-year period from 1930 to 2000, the last trajectories starting when the women of the first birth cohort were 20 years old. One cannot examine women's work histories over such a long period without referring to context. Women's labour market activity depends largely on the French context of work-family relationship, the pattern of which changed radically over this period as a combined result of changes in the labour market and transitions in family dynamics.

On average, French women in these birth cohorts, ever since the 1942 cohort, were better qualified than the men of the same generations (Léridon and Toulemon, 1995). This did not impact in a straightforward way on their labour market situation but it profoundly changed their aspirations and particularly affected changes in the family after the war (Barrère-Maurisson, 1984). Thus the background to the individual histories we want to characterise is structured around some major transitions. In France, this was a period of industrial restructuring, with the tertiary sector growing fast. Two landmark moments stand out: the massive entry of women into employment at the end of World War II and the economic crisis of 1974 which halted post-war growth. Further, France was unique in the scale and speed of feminisation of the workforce, and also in the fact that this was achieved through full-time work. The mean activity rate (between 29 and 59 years old) of women born between 1930 and 1950 grew from approximately 50% to more than 65% the more qualified born 1940 having a rate of 70% compared to 50% for those less qualified (Afsa Essafi and Buffeteau, 2006). Part-time working began to expand only in the early 1980s, as a result of the crisis and of policies that strongly encouraged it (Maruani, 2000).

This translates in our sample of respondents in the *Biographies et entourage* survey by visible changing trends in the proportion of women continuously employed: scarcely more than 5% for those born between 1930 and 1939 but close to 50% for those born between 1946 and 1950. The proportion of never-employed women is 4% for women born between 1930 and 1945 and just 2% for those born after 1945, with nearly all women having had at least one paid job in their lives. And finally, there are sharp differences between the birth cohorts, with the number of career interruptions falling very significantly over time.

On the family front, in contrast to earlier times, the distinction between domestic labour and employment became irreversible. In the post-war dual-earner family, the woman's earnings at

first constituted a top-up income (Barrère-Maurisson, 1992). The 1960s brought couples access to effective birth control and choice in conjugal matters. Fertility and marriage rates fell, divorce rates rose. In the early 1980s, dual-earner families became more common than those in which only the man was employed.

These generations were also pioneers in terms of demographic behaviour. Age at first marriage declined between the 1930 and 1950 birth cohorts, from about 23 to 22, while the period indicators for first marriage<sup>5</sup> rose steadily from 1950 to the mid-sixties. Mean age at childbirth declined steadily, from 28.7 for those born in 1913 to 26 for those born in 1945.

Women born between 1930 and 1950 were those who established the two-child family model as standard. In their mothers' birth cohorts (most of whom were born between 1900 and 1928) only about 28% of women had just two children, while the figure for women born between 1947 and 1952 is 44%. This is the combined result of reduced childlessness and a drop in fertility from the exceptionally high levels of the 1915-1935 cohorts (Daguet, 2000). Divorce rates, meanwhile, have been rising constantly since the 1960s, from 9.5 per 100 marriages in 1960 (Blayo, 1973) to 30.5 in 1985 and 38.2 in 2000 (Prioux *et al.*, 2010).

#### 4. A typological approach of work histories

Our objective is to explore the full complexity of the women's careers over the long term, including the timing of career interruptions, how long they lasted or the type of employment (full-time, part-time etc.). This can only be done by taking into account the trajectory as a whole, to draw out its main features and describe it synthetically. More precisely, we hope to pinpoint the regularities in the work histories of the population studied, identifying a small number of trajectory types. To do this, the first step is to code the work histories of the women respondents in the *Biographies et entourage* survey. These can be summarized as sequences of employment statuses year by year between the ages of 14 (end of mandatory schooling for the cohorts studied) and 50 (the age of youngest respondents at the time of the survey). This gives a set of 1487 sequences, all lasting 37 years, and each state having 4 modalities representing the four employment statuses, i.e. in education, inactive, employed part-time, employed full-time. This exercise reveals a huge variety of trajectories, with a thousand distinct sequences among the 1487 in the sample. To explore and describe these women's work histories therefore requires appropriate tools. "Holistic" methods for

---

<sup>5</sup> The sum of age-specific marriage rates for singles.

describing trajectories are suited to the task (Billari, 2005). They can be used to explore longitudinal data, reduce complexity by building synthetic typologies uncovering interesting patterns. Most of the methods currently used in the social sciences are either related to correspondence analysis methods (Grelet, 2002) or to sequence analysis methods (Abbott and Tsay, 2000)<sup>6</sup>. All have particularities, advantages and drawbacks that make them more or less suitable for the data studied<sup>7</sup> and the types of patterns one wants to identify, although they usually give quite similar results (Robette and Thibault, 2008). The method chosen here is Optimal Matching Analysis (OMA), the most widespread sequence analysis method in the social sciences. It is flexible enough to be readily adapted to the research subject. It also has the major conceptual advantage of jointly addressing the different temporal aspects of a sequence: the moment of a transition, the duration of a stage and the order of the stages. Most of the conventional methods do not do this.

Optimal Matching Analysis is based on a set of dynamic algorithms derived mainly from molecular biology. It was introduced into the social sciences by Andrew Abbott in the 1980s (Abbott and Forrest, 1986). The principle is based on the notion of similarity between pairs of sequences. The dissimilarity between two sequences is measured in terms of the "cost" of transforming one into the other. The transformation is performed using three types of basic operation: insertion (inserting an element into the sequence), deletion (deleting an element) and substitution (replacing one element by another). A cost is associated with each operation; the distance between two sequences is thus defined as the minimum cost of the operations required to transform one into the other. Matching the entire set of sequences creates a matrix of pairwise distances which is then used to group together those that are most similar, e.g. using clustering techniques, and so obtain a typology<sup>8</sup>.

The choice of costs is an important in the OMA process (Lesnard, 2010). In particular, how one arbitrates between levels of insertion/deletion<sup>9</sup> costs and of substitution costs will determine whether more importance is given to regularities in the order of elements or in the moment of transition. For our study, to avoid imposing any presupposed hierarchy on occupational situations, substitution cost was set at the same value regardless of the elements

---

<sup>6</sup> For a comprehensive review of these two sets of methods, see Robette, 2011.

<sup>7</sup> For instance, some methods are more sensitive to the ordering of sequences, while some others are more sensitive to rare events.

<sup>8</sup> For a detailed presentation of the different stages of a sequence analysis by Optimal Matching see for example MacIndoe and Abbott (2004).

<sup>9</sup> Also called *indel*, a contraction of "insertion" and "deletion".

replaced. Furthermore, we need to distinguish sequences according to age at labour market entry which is a decisive marker in a woman's occupational history mainly because it corresponds to the end of education (in 71% of cases) and distinguishes between education levels. In other words, timing has to be underlined and we have to prevent too many uses of *indel* operations. The *indel* cost is therefore set at quite a high level, slightly over half the maximum substitution cost (see substitution and *indel* costs in Appendix A). These analyses were computed using TraMineR package in R software (Gabadinho *et al*, 2011).

## 5. Seven types of female work histories

The matrix of distances between the 1487 sequences derived from the Optimal Matching algorithm was subjected to hierarchical clustering using the Ward aggregation criterion. A division into seven clusters was chosen; this accounts for 80% of the information (i.e. of the variety of trajectories). The number of clusters was chosen to obtain a set of homogeneous and theoretically relevant groups that would sufficiently reflect the variety within the sample of trajectories and facilitate description.

The seven clusters obtained reflect a partition of work histories into women's experience comprising 'uninterrupted full-time employment' by far the most common work career (57.1%, clusters 1 and 2) in this representative sample of women born between 1930 and 1950, and 'never having been employed or only just after school' (17.9%, cluster 3) (Table 2).

*Table 2. Clusters of female occupational trajectories*

Cluster	Main features of employment history	N	%
1	always full-time, started early (before age 20)	555	37.3
2	always full-time, started late (after age 20)	294	19.8
3	never employed or stopped early (before age 30)	266	17.9
4	interruption (from age 20-25 to 30-40)	181	12.2
5	switch to part-time (after age 30)	85	5.7
6	stopped work late (after age 35)	59	4.0
7	interruption (c. age 25-30), re-entry part-time (c. age 30-40)	47	3.2
Total		1487	100.0

Source: *Biographies et entourage* (2000)

Reference population: Female survey respondents born between 1930 and 1950

The first cluster – by far the largest, with over a third of the respondents – is that of women who started work before the age of 20 and continued to work full-time throughout their careers. These women were often poorly qualified and many were clerical or sales staff (see Appendix B). The next group (19.8%) consists of women who were also in full-time employment throughout their careers but started work later than the first group, after the age of 20. Having had time to continue their education they were by far the best qualified (over half had a diploma equivalent to two years in higher education) and nearly 40% were in higher-level occupations.

All the other clusters (43% of total) consist of women who had at some point worked part-time or been inactive. Women in cluster 3 have never been employed or have stopped work for good before age 30, while the women in cluster 6 (4%) had quit the labour market later, after the age of 35. The women in cluster 4 (12.2%) interrupted their careers, stopping work at 20-25 years old and resuming around age 30-40. Those of cluster 7 (3.2%) stopped work at age 25-30 and resumed work part-time at 30 to 40 years old. Cluster 5 women (5.7%) shifted to part-time work after the age of 30.

*Table 3. Trajectory types by birth cohort*

Cluster	Labels	Year of birth		
		1930-1939	1940-1945	1946-1950
1	always full-time, started early	35.2	38.2	38.5
2	always full-time, started late	13.7	17.9	21.9
3	never employed or stopped early	24.0	17.4	12.3
4	interruption	10.7	14.5	7.7
5	switch to part-time	3.9	2.7	6.1
6	stopped work late	8.4	4.1	6.4
7	Interruption, re-entry part-time	4.1	5.3	7.1
Total		100.0	100.0	100.0

Source: *Biographies et entourage* (2000)

Reference population: Female survey respondents born between 1930 and 1950

As expected, the distribution of the trajectory types by year of birth shows that the patterns of women's labour force involvement over the course of their lives have changed over time (Table 3). The main result has been a sharp increase in the proportion of women entering the labour force late and employed full-time (cluster 2) and the decline in those who were never or hardly ever in employment (cluster 3). This results from women's increased levels of qualification and the gradual demise of the family model with breadwinner husband and wife at home. Women who started work early and remained in continuous full-time employment

(cluster 1) account for nearly two out of five women's careers, varying little between birth cohorts. The fluctuations for clusters 4 to 7 should be treated with caution given the small numbers of women concerned.

Occupational and family trajectories interact in complex, dynamic ways (see for example Pailhé and Solaz, 2009). Without trying to model this process in depth, we can illustrate it by correlating our typology with a few indicators that synthesise particular aspects of family trajectories such as marriage and fertility. For example, unsurprisingly, women who were employed throughout their careers (clusters 1 and 2) had relatively few children while those who stopped early or interrupted their careers (clusters 3, 4 and 7) had more (see Appendix B). Women who started late, after higher education, and were continuously employed (cluster 2) were more likely than others to go through a time of separation or divorce (see Appendix B). Conversely, those who were never employed or stopped work early (cluster 3) had the most stable conjugal trajectories.

*Table 4. Percentage of respondents in each trajectory type having inactive mothers*

Cluster	Labels	Inactive mother		Total
		yes	no	
1	always full-time, started early	44.2	55.8	100.0
2	always full-time, started late	60.3	39.7	100.0
3	never employed or stopped early	64.8	35.2	100.0
4	interruption	53.8	46.2	100.0
5	switch to part-time	38.5	61.5	100.0
6	stopped work late	50.0	50.0	100.0
7	Interruption, re-entry part-time	55.4	44.6	100.0
Total		52.5	47.5	100.0

Source: *Biographies et entourage* (2000)

Reference population: Female survey respondents born between 1930 and 1950

The mother's economic activity pattern is also a determining factor in the differentiation of women's work histories. Table 4, for example, shows that relatively few respondents who were always employed and started early (cluster 1) had mothers who stayed at home compared to respondents who started later (cluster 2) and those who were economically inactive for most of their careers (cluster 3). However, this is still a relatively crude indicator and does not take account of the diversity of the mothers' trajectories. To examine this in greater detail and better understand the link between the mothers' work histories and those of their daughters, we also take a typological approach to describe the mothers' trajectories.

## 6. Work histories of mothers of women born between 1930 and 1950 in France

The respondents' mothers were born between 1886 and 1935; half were born between 1907 and 1920. The lifetime fertility of the mothers in these birth cohorts was 3 children (2.8 per mother born in 1900, with the maximum observed to date of 3.2 for those born in 1926).

Studying the respondents' mothers' work histories obviously selects women in these birth cohorts who had children and excludes all those who did not – i.e. a quarter of women born in 1900, falling to 19% of those born in 1920 (Daguet, 2000). These childless women from these generations born in the first decades of the XX<sup>th</sup> century certainly played an important part in the development of women's labour force participation but can obviously not be studied with our data.

Our respondents' mothers, while having heavy family duties, also played a significant part in the transformation of industrial wage labour, in the general context of the agricultural crisis of the late 19th century, the start of urban expansion and intensive industrial development which happen later in France compared to England (Battagliola, 2000).

As with the respondents, we constructed a typology of the mothers' work histories<sup>10</sup>.

However, only three employment statuses can be distinguished<sup>11</sup>: in education, employed and non-employed, with no distinction between full-time and part-time.

*Table 5. Typology of respondents' mothers' trajectories*

Cluster	Main features of employment history	N	%
M <sub>1</sub>	Always employed	487	32.8
M <sub>2</sub>	Stopped work after age 20	370	24.9
M <sub>3</sub>	Never employed	347	23.3
M <sub>4</sub>	Interruption (stopped after age 18, resumed before age 40)	283	19.0
	Total	1487	100.0%

Source: *Biographies et entourage* (2000)  
Reference population: the 1487 respondents' mothers

This gives a four-cluster distribution (Table 5) explaining 80% of mother careers' variability.

The largest cluster is that of mothers who were always employed (32.8%). Next come mothers who stopped work after age 20 (24.9%) and those who were never employed

<sup>10</sup> Once again we used Optimal Matching with fixed substitution costs and an *indel* cost slightly greater than half the substitution cost (see Appendix C).

<sup>11</sup> A fourth modality was added, for years on which there were no data. This has the advantage of enabling all the mothers to be included in the analysis. In fact the years with no data were relatively few, and mostly corresponded to short periods in fairly chaotic trajectories. Our explorations showed that the missing values had a negligible impact on the construction of the typology.

(23.3%). Last comes a cluster of mothers with interrupted careers, stopping work after age 18 and resuming it before age 40 (19.0%).

Comparing the broad features of the respondents' careers with those of their mothers requires caution, as the typologies are not built from strictly identical information, still it is informative. Continuous (or nearly continuous) employment is naturally markedly more common among the respondents. The total of always-employed women, regardless of the age at which they joined the labour force, is over 60% for the respondents but little more than a third for their mothers. Conversely, about a fifth of the respondents was never employed or stopped work early, compared to nearly half of their mothers. This is perfectly consistent with the macro trend observed for employment over the generations of women in France.

## 7. Inter-generational mother-daughter trajectories

Our aim now is not to identify the influence of the mother's work history on that of her daughter, but to see whether we can identify preferential dyads among the lineages – recurrent sequences conditioned by the social cohesion between mother and daughter: an internal *structure* in the lineage concerning work patterns.

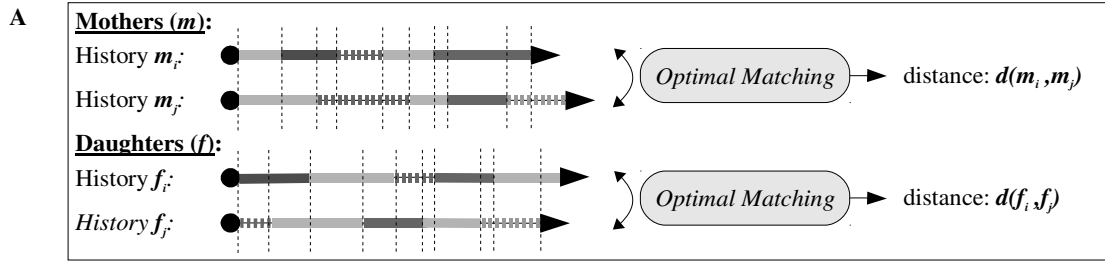
To pursue the analysis by studying mother-daughter pairs of trajectories, simply correlating the typologies of daughters' work histories with those of their mothers as determined above (Tables 2 and 5), produces a contingency table of  $7 \times 4 = 28$ . Still a 28-cluster distribution is hard to describe. Moreover it would be based on the specific characteristics of the daughter's (respectively the mother's) individual work histories, whose correlations would then be identified *ex-post*: the characteristics underlying mother-and-daughter dyads would be hidden. We need to use a method that specifically focuses on the linked characteristics of the two work histories.

The typological approach used so far lends itself well to mother-daughter trajectory sequences, provided that the work histories of mothers and daughters are paired. To do so, several options are open to us. First, we could concatenate a daughter's trajectory to her mother's, to form a single sequence. Still those work careers do not develop in time in strict chronological succession, they overlap. Then we could construct combined states from the available dated information, i.e. join the employment positions of mother and daughter at the precise moment they coincide in time. Nevertheless this way only parts of sequences would concern both mothers and daughters and the resulting number of states would dilute the issue. And above all, this would infer a direct relation between mothers' and daughters' employment

statuses in a transversal way, i.e. at a given moment, which is meaningless from our point. Indeed the daughter's situation at a given moment is unlikely to be determined by her mother's situation at the same moment, but rather by the shared part of background that conditions both work histories. This type of long-term correlation is too complex a process for this kind of descriptive work. Practically, this means our data cannot be assimilated to multidimensional sequences<sup>12</sup>. Our approach is to build a typology of pairs of mother-daughter work histories that captures in detail the correspondence between some patterns of the daughter's work history and some of her mother's.. This approach, "Dyadic Sequence Analysis" (DSA), breaks down into the several stages described in the next section.

## 8. Constructing a typology of mother-daughter trajectories

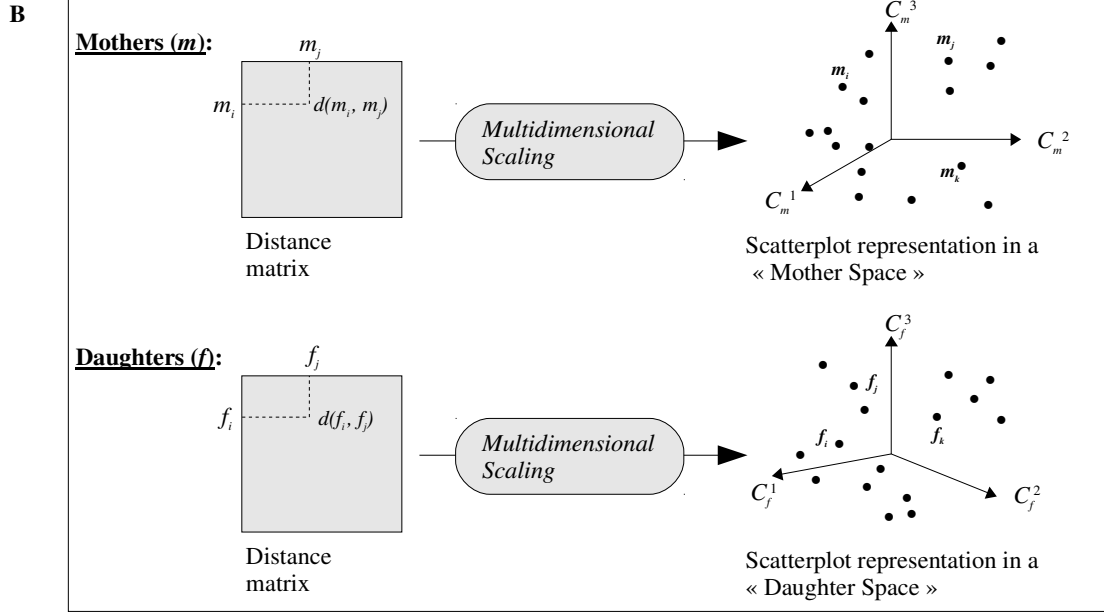
Stage A:



At this stage, Optimal Matching (OM) is used to calculate informational distances associated with each pair of mother's and daughter's trajectories. This stage gives us two symmetric distance matrices  $\mathbf{M} = (m_{ij})_{i,j}$  where  $m_{ij} = d(m_i, m_j) = d(m_j, m_i)$  (respectively  $\mathbf{F} = (f_{ij})_{i,j}$  where  $f_{ij} = d(f_i, f_j) = d(f_j, f_i)$ ) in which the diagonal is zero. A brief presentation of the OM algorithm may be found in a previous section of this paper (for a more detailed presentation, see for instance MacIndoe and Abbott, 2004).

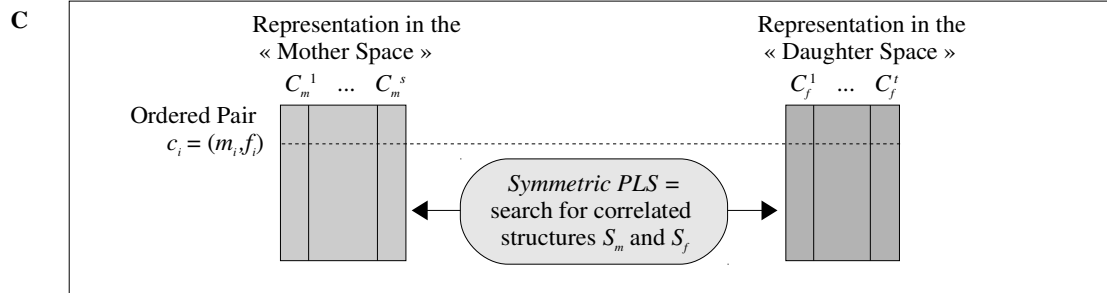
<sup>12</sup> The issue of uncovering patterns of multidimensional sequences has already been explored widely; see for instance Pollock (2007), Robette (2010) or Gauthier *et al* (2010).

## Stage B:



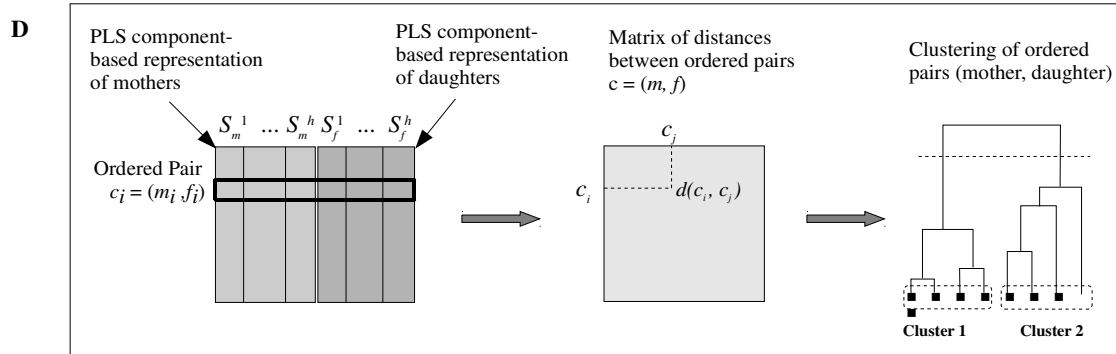
The matrix  $M$  (respectively  $F$ ) of distances between the mothers' trajectories (and respectively the daughters') is converted to a spatial representation by multidimensional scaling (MDS) (Kruskal and Wish, 1984; see Halpin and Chan, 1998 and Piccarreta and Lior, 2010 for MDS applied to sequence analysis), i.e. represented as a scatterplot of points in a multidimensional "mothers' space" (resp. "daughters' space") with respect to a system of principle components  $C_m^1, C_m^2$ , etc. (resp.  $C_f^1, C_f^2$ , etc.). For a quick presentation of MDS, see appendix D. The sequence of components  $C_m^1, C_m^2$ , etc. (resp.  $C_f^1, C_f^2$ , etc.) provides a hierarchical breakdown of the heterogeneity of the mothers' (resp. the daughters') trajectories, in the sense that each component complements the preceding ones in an optimal manner. At the end of this stage, each mother-daughter pair is described both by the coordinates of the mother's trajectory in the mothers' space and those of the daughter's trajectory in the daughters' space.

## Stage C:



We next look for structural links between the mothers' and daughters' trajectories. The symmetric (or canonical) PLS method (Partial Least Square) (Bry, 1996; de Jong *et al*, 2001) seeks structures *common* to the variability of the mothers' trajectories and that of the daughters' trajectories. These structures are extracted under the form of components denoted  $S_m^k$  (respectively  $S_f^k$ ) for the mothers (resp. daughters). For a quick presentation of symmetric PLS, see appendix E.

*Stage D:*



The Euclidean coding of the mothers' and daughters' trajectories, restricted to "common" components  $S_m^k$  and  $S_f^k$ , finally provides a base for clustering the mother-daughter pairs. We use Hierarchical Clustering Analysis with Ward criterion. This gives a typology of pairs based on structures common to the discrepancies of the mothers' histories and to those of their daughters'.

One advantage of the DSA approach is that the mothers' and daughters' trajectories need not take the same form. Their work histories can differ in length and in their set of possible states. Here the trajectories considered for mothers and daughters both go from age 14 to age 50, but those of the daughters also distinguish between full-time and part-time employment, to make the most of the information provided by the survey.

The typology of mother-daughter dyads resulting from this approach comprises nine clusters, as shown in Table 6.

We have sorted the resulting typology according to mothers' employment profile (i.e. from. always employed to never employed), and secondarily according to daughters' profile.

*Table 6: Typology of mother-daughter trajectories*

Cluster	Main features of the dyads		N	%	
	mothers	daughters			
1	always employed	always employed	346	23.3%	<b>38.3</b>
2	always employed	full-time to part-time	75	5.0%	
3	always employed	alternating employed / non-employed	148	10.0%	
4	interruption	always employed	147	9.9%	<b>33.6</b>
5	alternating employed / non-employed	stopped work early (before age 30)	101	6.8%	
6	stopped work before age 35	always employed	251	16.9%	
7	never employed	always employed	151	10.2%	<b>28.2</b>
8	never employed	long period of part-time	82	5.5%	
9	never employed	never employed or long period non-employed	186	12.5%	
<i>Total</i>			1487	100%	100%

Source: *Biographies et entourage* (2000)

Reference population: the 1487 female respondents and their mothers

The largest cluster is that in which both mothers and daughters were always held a paid job (23.3%, cluster 1), as naturally results from the main situations identified earlier. The respondents (i.e. daughters) in this cluster are relatively more often only daughters, born in France between 1946 and 1950 (i.e. in the youngest cohorts), with few qualifications, working mainly in clerical or sales jobs or intermediate occupations. So these are lineages of working class and middle class women, whose employment was and is essential to the family economy. In France, unlike the UK or Germany, these lineages of working women helped to establish women's continuous presence in the labour market early on, undermining the male-bread-winner-and-housewife model of the family. There are two other clusters with mothers always in employment: in one, the daughters shift from full-time employment to part-time while raising children (cluster 2, 5%) and in the other the daughters alternate between work and home (cluster 3, 10%).

In clusters 4 and 5, mothers have experienced several transitions between employment and unemployment. In the former (9.9%), they have only interrupted their career for a given period and their daughters have a complete career. Conversely, in the latter (6.8%), mothers often alternate between labour market and inactivity and their daughters tend to stop working before 30.

In cluster 6, which is the second largest (16.9%), the mothers began their trajectory in employment but stopped before the age of 35, while the daughters were employed throughout.

Finally, there are several clusters in which the mothers were never employed. In one, the daughters in turn were never employed, or only briefly (cluster 9, 12.5%); many of these daughters were born abroad, came from large families and had several children and few qualifications. In another, the daughters were economically active but their careers included a long period of part-time working (cluster 8, 5.5%); in the last, the daughters worked full-time throughout (cluster 7, 10.2%). The latter group is mostly composed of highly qualified women, both they and their husbands working in higher-level occupations. These three lineage profiles suggest a transition within the middle and upper classes where two models co-exist: a "traditional" model and one where there is investment in the daughters' education and the women are in higher-level occupations.

If we now put the focus on the respondent's (daughter's) employment profile, we find that 60% were always in employment (clusters 1, 4, 6 and 7). Still their mothers show various activity patterns: some respondents are daughters of working-class lineages where the mother too was always employed (cluster 1), or lineages where the mother was never employed, or stopped work when their family started, or interrupted her career but started again (resp. 6, 7 and 4), and where the daughters, often highly qualified and in high-level occupations, have relatively few children. In the lineage profiles of non-employed respondents, most of the mothers were never employed (cluster 9) or alternated between employment and home, with the daughters stopping work definitively before age 30 (cluster 5).

The question remains of how to characterise the clusters obtained, in both methodological and conceptual terms. A trajectory type cannot be explained by its end point or a life course by its final state. We are confronted with problems of endogeneity, anteriority and causality. So, unsurprisingly, descriptive methods have to give way to other tools better equipped to help us pursue the analysis of social transformation suggested by and contained in these typologies of trajectories. In this view, DSA proves more powerful than simple contingency tables (combining two independent typologies) to uncover the patterns of mother-and-daughter dyads of careers.

## 9. Discussion

Using the rich data from the *Biographies et entourage* survey (INED, 2000), from which one can track respondents' work histories for 37 years (from age 14 to 50) and also those of their mothers, we set out to make a typological analysis of women's work histories using a sequence analysis approach (Optimal Matching).

With this approach we were able to observe, for example, that the proportion of economically active, poorly qualified women remained fairly stable across the birth cohorts from 1930 to 1950, while the proportion of women starting work after higher education and thereafter continuously in employment increased. Never-employed profiles and interrupted careers became fewer and fewer.

The second strand examined in this paper was the inter-generational pattern of women's occupational profiles within lineages, pairing the work histories of mothers and daughters. The first results presented here are promising and open new perspectives for studying long term trends and understanding specific intra-family features and continuities that are worth investigating further. The typologies obtained shed new light on transmission, leaving aside a basic determinism and showing the relative multiplicity of career pathways open to children starting from the same parental trajectory: some never employed mothers have daughters with incomplete careers, but others' daughters work always full-time, and so on. The differences depend on characteristics such as social class, educational background, cohort or rank among siblings.

From a methodological point of view, DSA provides a flexible way to uncover patterns of dyads of sequences, with very few constraints about the data: sequences within the dyads do not have to be contemporaneous, nor to be of the same length or even of the same nature (e.g. to have the same state space). And yet, combining OM with euclidean tools, DSA enables a straightforward and computationally efficient multidimensional approach to trajectories' pattern mining. The range of potential applications for DSA thus largely exceeds intergenerational social mobility studies.

## References

- Abbott, A. and Forrest, J. (1986) Optimal Matching Methods for Historical Sequences. *Journal of Interdisciplinary History*, **16**, 471-494.
- Abbott, A. and Tsay, A. (2000) Sequence Analysis and Optimal Matching Methods in Sociology. *Sociological Methods and Research*, **29**(1), 3-33.
- Afsa Essafi, C., Buffeteau, S. (2006) L'activité féminine en France : quelles évolutions récentes, quelles tendances pour l'avenir ?. *Economie et statistiques*, **398-399**, 85-97.
- Antoine, P., Bry, X. and Diouf P. D. (1987) La fiche « Ageven » : un outil pour la collecte des données rétrospectives, *Techniques d'Enquête* (CAN), **13**(2), 173-181.
- Barrère-Maurisson, M-A. (1984) Du travail des femmes au partage du travail. Une approche des régulations familiales face aux évolutions du travail : le cas de la France depuis 1945. *Sociologie du travail*, **3**, 243-254.
- Barrère-Maurisson, M-A. (1992) *La division familiale du travail, la vie en double*. Paris : PUF.
- Battagliola, F. (2000) *Histoire du travail des femmes*. Paris : La Découverte.
- Beller, E. (2009) Bringing Intergenerational Social Mobility Research into the Twenty-first Century: Why Mothers Matter. *American sociological review*, **74**(4), 507-528.
- Billari, F. C. (2005) Life course analysis: two (complementary) cultures? Some reflections with examples from the analysis of the transition to adulthood. *Advances in life course research*, **10**, 261-281.
- Blayo, C. (1973) Evolution des désunions en France depuis 1950. *Population*, **3**, 604-619.
- Bry, X. (1996) *Analyses factorielles multiples*. Paris : Economica Poche.
- Brzinsky-Fay, C. and Kohler, U. (2010) New Developments in Sequence Analysis. *Sociological Methods & Research*, **38**(3), 359-364.
- Daguet, F. (2000) L'évolution de la fécondité des générations nées de 1917 à 1949 : analyse par rang de naissance et niveau de diplôme. *Population*, **55**, 1021-1034.
- De Jong, S., Wise, B.M. and Ricker, N.L. (2001) Canonical partial least squares and continuum power regression. *Journal of Chemometrics*, **15**, 85-100.
- Gabadinho, A., Ritschard, G., Müller, N.S. and Studer, M. (2011) Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software*, **40**, 1-37.
- Gauthier, J., Widmer, E. D., Bucher, P. and Notredame, C. (2010) Multichannel sequence analysis applied to social science data. *Sociological Methodology*, **40**, 1-38.
- GRAB (2009) *Biographies d'enquêtes : bilan de 14 collectes biographiques*. Paris : INED. [[http://grab.site.ined.fr/fr/editions\\_en\\_ligne/biographies\\_enquetes/](http://grab.site.ined.fr/fr/editions_en_ligne/biographies_enquetes/)]
- GRAB (2009) *Fuzzy States: observing, modelling and interpreting complex trajectories in life histories*. Paris: INED/CEPED.
- Grelet, Y. (2002) Des typologies de parcours. Méthodes et usages. *Document Génération* **92**, **20**.
- Halpin, B. and Chan, T.W. (1998) Class careers as sequences: an optimal matching analysis of work-life histories. *Eur. Sociol. Rev.*, **14**, 111-130.

- Kruskal, J. B. and Wish, M. (1984) *Multidimensional Scaling*. Beverly Hills: Sage.
- Lelièvre, E. and Vivier, G. (2001) Evaluation d'une collecte à la croisée du quantitatif et du qualitatif : l'enquête Biographies et entourage. *Population*, **6**, 1043-1073.
- Léridon, H. and Toulemon, L. (1995) Recent changes in the process of family formation in France: a cohort perspective. In *The New Role of Women, Family formation in modern societies* (ed P.Blossfeld), the Social Equality series of Westview Press.
- Lesnard, L. (2010) Setting Cost in Optimal Matching to Uncover Contemporaneous Socio-Temporal Patterns. *Sociological Methods & Research*, **38**, 389-419.
- MacIndoe, H. and Abbott, A. (2004) Sequence analysis and optimal matching techniques for social science data. In *Handbook of Data Analysis* (eds M. Hardy and A. Bryman), London: Sage.
- Maruani, M. (2000) *Travail et emploi des femmes*. Paris : La Découverte.
- Piccarreta, R. and Lior, O. (2010) Exploring sequences: a graphical tool based on multi-dimensional scaling. *J. R. Statist. Soc. A*, **173**, 165-184.
- Prioux, F., Mazuy, M. and Barbieri, M. (2010) Recent Demographic Developments in France: Fewer Adults Live with a Partner. *Population-E*, **65** (3), 363-414.
- Pollock, G. (2007) Holistic trajectories: a study of combined employment, housing and family careers by using multiple-sequence analysis. *J. R. Statist. Soc. A*, **170**, 167-183.
- Robette, N. (2010) The diversity of pathways to adulthood in France: evidence from a holistic approach. *Advances in life course research*, **15**, 89-96.
- Robette, N. (2011) *Explorer et décrire les parcours de vie: les typologies de trajectoires*. Paris : CEPED, 86 p.
- Robette, N. and Thibault, N. (2008) Comparing qualitative harmonic analysis and optimal matching. An exploratory study of occupational trajectories. *Population-E*, **64**, 621-646.
- Vallet, L.A. (2001) Forty years of social mobility in France. Change in social fluidity in the light of recent models. *Revue Française de Sociologie*, **42**(suppl.), 5-64.

Appendix A: Substitution and *indel* costs for the respondents

Substitution costs

	<i>education</i>	<i>non- employed</i>	<i>employed part-time</i>	<i>employed full-time</i>
<i>education</i>	0	2	2	2
<i>non-employed</i>	2	0	2	2
<i>employed part-time</i>	2	2	0	2
<i>employed full-time</i>	2	2	2	0

*indel* cost = 1.1

## Appendix B: Characteristics of respondent's employment histories by cluster

cluster	qualification				Total
	none	<bac	bac	>bac	
1	6,3	60,9	20,0	12,8	100,0
2	5,8	19,0	19,4	55,8	100,0
3	18,8	47,4	20,3	13,5	100,0
4	11,6	60,8	15,5	12,1	100,0
5	11,8	36,5	15,3	36,5	100,0
6	13,6	47,5	20,3	18,6	100,0
7	21,3	42,5	17,0	19,2	100,0
<b>Total</b>	<i>10,2</i>	<i>47,7</i>	<i>19,0</i>	<i>23,1</i>	<i>100,0</i>

cluster	ego's last SOC						Total
	non-empl. <sup>13</sup>	self-empl.	intermediate occ.	higher-level occ.	clerical and sales	manual	
1	8,1	2,2	12,6	32,4	41,8	2,9	100,0
2	9,5	1,7	38,1	28,2	20,4	2,1	100,0
3	72,9	1,1	2,6	4,9	16,6	1,9	100,0
4	13,8	1,1	9,9	21,6	48,1	5,5	100,0
5	16,5	3,5	27,1	20,0	31,8	1,2	100,0
6	42,4	3,4	6,8	15,2	25,4	6,8	100,0
7	17,0	0,0	8,5	21,3	48,9	4,3	100,0
<b>Total</b>	<i>22,8</i>	<i>1,8</i>	<i>16,0</i>	<i>23,6</i>	<i>32,8</i>	<i>3,0</i>	<i>100,0</i>

cluster	number of children				Total
	0	1	2	3 or +	
1	18,2	25,6	37,8	18,4	100,0
2	21,4	23,8	33,7	21,1	100,0
3	0,8	10,1	34,2	54,9	100,0
4	2,8	7,7	39,8	49,7	100,0
5	7,1	14,1	50,6	28,2	100,0
6	10,2	22,0	30,5	37,3	100,0
7	0,0	4,3	31,9	63,8	100,0
<b>Total</b>	<i>12,3</i>	<i>18,8</i>	<i>36,9</i>	<i>32,0</i>	<i>100,0</i>

cluster	at least 1 separation before age 50		Total
	yes	no	
1	23,8	76,2	100,0
2	26,9	73,1	100,0
3	9,0	91,0	100,0
4	21,0	79,0	100,0
5	16,5	83,5	100,0
6	25,4	74,6	100,0
7	4,3	95,7	100,0
<b>Total</b>	<b>20,4</b>	<b>79,6</b>	<b>100,0</b>

Reference population: the 1487 women respondents in the *Biographies et entourage* survey

## Appendix C: Substitution and *indel* costs for respondents' mothers

### Substitution costs

	<i>education</i>	<i>non-employed</i>	<i>employed</i>	<i>not specified</i>
<i>education</i>	0	2	2	0
<i>non-employed</i>	2	0	2	0
<i>employed</i>	2	2	0	0
<i>not specified</i>	0	0	0	0

*indel* cost = 1.1

## Appendix D: Multidimensional Scaling (MDS)

Given the matrix of distances  $\|\xi_i - \xi_j\|$  between  $n$  points  $\{\xi_i; i=1, n\}$  in a euclidean space, MDS allows rebuild the image of the unit scatterplot in the basis of its principal components:

■ *Finding the scalar product matrix of vectors centered on their centroid:*

Centering vectors on their centroid: let  $\bar{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_i$  and  $x_i \stackrel{\text{def.}}{=} \xi_i - \bar{\xi} \quad i=1, n$ .

Then:

$$\forall i, j=1, n : \|\xi_i - \xi_j\|^2 = \|x_i - x_j\|^2 = \|x_i\|^2 + \|x_j\|^2 - 2 \langle x_i, x_j \rangle$$

$$\langle x_i, x_j \rangle = \frac{1}{2} (\|x_i\|^2 + \|x_j\|^2 - \|x_i - x_j\|^2) \quad (1)$$

Besides, in view of the Koenig equality applied to  $\{x_i; i=1, n\}$  with centroid  $\bar{x} = 0$ :

$$\forall i=1, n : \sum_{j=1}^n \frac{1}{n} \|x_i - x_j\|^2 = \|x_i - 0\|^2 + \sum_{j=1}^n \frac{1}{n} \|x_j - 0\|^2 \quad (2_i)$$

$$\text{Let } D_0 = \sum_{j=1}^n \frac{1}{n} \|x_j\|^2 \text{ and } D_i = \sum_{j=1}^n \frac{1}{n} \|x_i - x_j\|^2.$$

From (2<sub>i</sub>), we have:

$$\forall i=1, n : \|x_i\|^2 = D_i - D_0 \quad (3)$$

Summing up equations (2<sub>i</sub>) over  $i$  and dividing by  $n$ , we get:

$$\frac{1}{n^2} \sum_{j=1}^n \|x_i - x_j\|^2 = 2 D_0 \quad \square \quad D_0 = \frac{1}{2n^2} \sum_{j=1}^n \|x_i - x_j\|^2 \quad (4)$$

So, from (4) and (3), draw every  $\|x_i\|^2$ , and then, from (1), every  $\langle x_i, x_j \rangle$ .

■ *Finding the principal components:*

Let  $S$  be the matrix  $(\langle x_i, x_j \rangle)_{i,j=1,n}$ . Let  $\lambda_k$  denote the  $k^{\text{th}}$  eigenvalue in decreasing order and  $v_k$  be the associated unit norm eigenvector. Then, the principal components  $f^k$  of  $\{x_i; i=1, n\}$  are:

$$f^k = \sqrt{\lambda_k} v_k$$

## Appendix E: Symmetric PLS

Given two data matrices  $X(n,p)$  and  $Y(n,q)$  containing respectively  $p$  and  $q$  numeric variables describing the same  $n$  statistical units, the purpose of symmetric PLS is to extract two sequences of uncorrelated components  $\{f^k, k=1, K\}$  and  $\{g^k, k=1, K\}$ , such that,  $\forall k$ :

- $f^k$  (respectively  $g^k$ ) belongs to the space spanned by  $X$ 's (resp.  $Y$ 's) columns;
- $f^k$  (respectively  $g^k$ ) captures as much as possible of  $X$ 's (resp.  $Y$ 's) variance unaccounted for by previous components;
- $f^k$  and  $g^k$  are as correlated as possible.

Such components are extracted through the following algorithm.

### Rank 1 components:

Let  $f^1 = Xu_1$  with  $\|u_1\| = 1$ ;  $g^1 = Yv_1$  with  $\|v_1\| = 1$

Vectors  $u^1$  and  $v^1$  are the solutions of the following program:

$$\mathbf{Q}(X,Y): \max_{\substack{u \in P, u'u=1 \\ v \in Q, v'v=1}} \text{cov}(f,g) \Leftrightarrow \max_{\substack{u \in P, u'u=1 \\ v \in Q, v'v=1}} \langle Xu | Yv \rangle_P \text{ where } P = \frac{1}{n}I$$

$$L = v'Y'PXu - \lambda(u'u - 1) - \mu(v'v - 1)$$

$$\square_u L = 0 \quad \square_u X'PYv = 2\lambda u \quad (1) ; \quad \square_v L = 0 \quad \square_v Y'PXu = 2\mu v \quad (1')$$

$u'(1)$  and  $v'(1')$  give:

$$u'X'PYv = 2\lambda u'u = 2\lambda ; \quad v'Y'PXu = 2\mu v'v = 2\mu$$

$$\square \text{cov}(f,g) = 2\lambda = 2\mu \stackrel{\text{def.}}{=} \eta$$

which implies that  $\eta$  be maximum.

Besides:

$$(1,1') \quad \square X'PYY'PXu = \eta u \quad (2) ; \quad Y'PXX'PYv = \eta v \quad (2')$$

So, the solution vector  $u$  (resp.  $v$ ) is the eigenvector characterized by (2) (resp. (2')) associated with the largest eigenvalue.

**Rank  $k > 1$  components:**

Rank  $k$  component  $f^k$  (resp.  $g^k$ ) must be uncorrelated to the former rank ones  $f^1, \dots, f^{k-1}$  (resp.  $g^1, \dots, g^{k-1}$ ). To ensure that, we define:

$$X_0 = X ; Y_0 = Y \quad \text{and} \quad \forall k > 1 : X_k = \Pi_{f^k} X_{k-1}, Y_k = \Pi_{g^k} Y_{k-1}$$

To put it more statistically,  $X_k$  (resp.  $Y_k$ ) is made of the residuals of  $X_{k-1}$  (resp.  $Y_{k-1}$ ) regressed on  $f^k$  (resp.  $g^k$ ). Then, we look for:

$$(u_k, v_k) = \text{sol. of } Q(X_{k-1}, Y_{k-1})$$