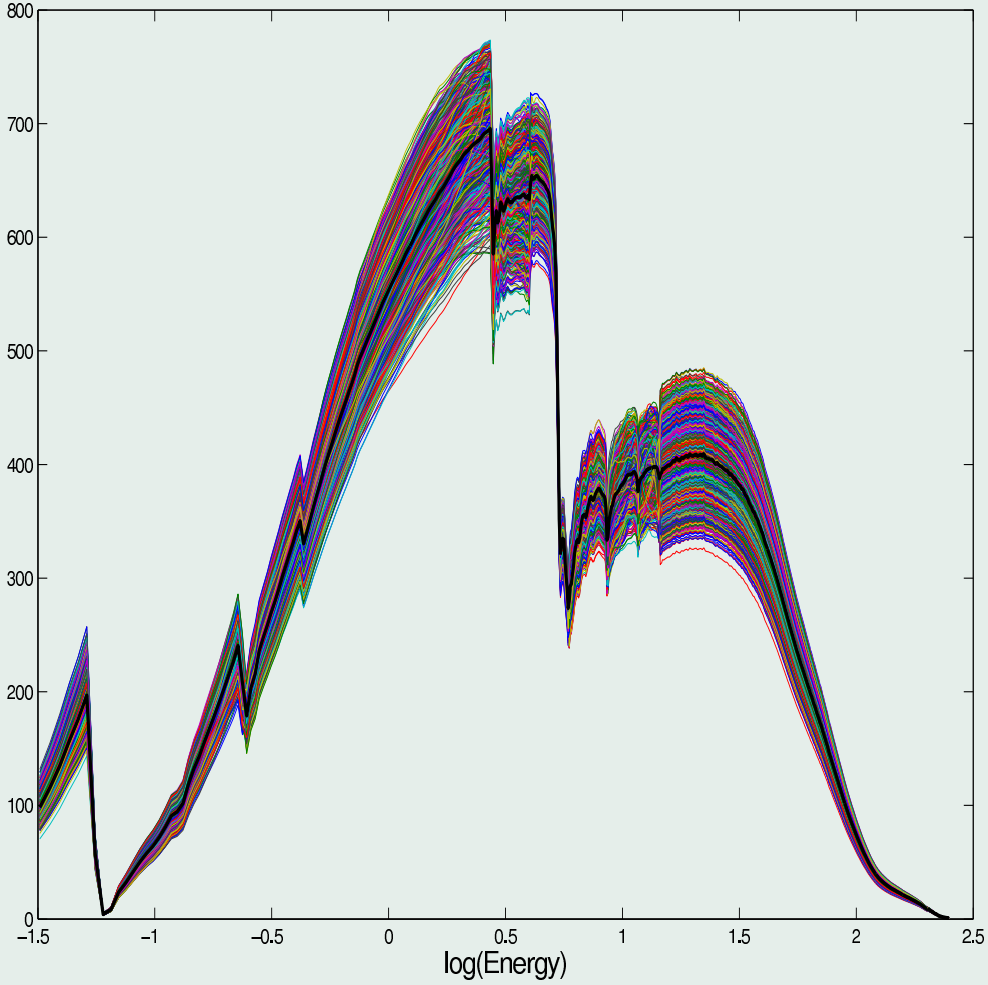
Dimensionality Reduction of Functional Data

Rima Izem
Statistics Department
Harvard University
izem@stat.harvard.edu



PCA

ICA

Functional Data

- Multivariate data: $f_i(x) = (f_i(x_1), \ldots, f_i(x_d))', 1 \leq i \leq n$
- In our example, dimension d is 1078 and sample size n is 1000.

Dimensionality reduction

- Use basis functions, polynomial, wavelets, splines ..etc

$$f_i(x) = \sum_{k=1}^{d'} \alpha_{i,k}\psi_k(x); d' < d$$

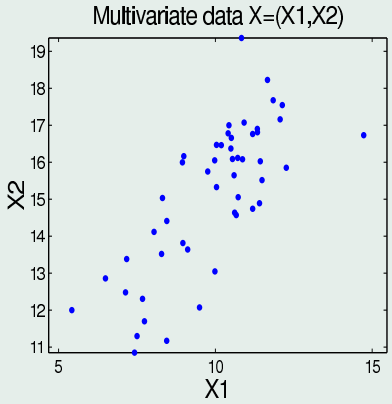
- Directions in feature space , decomposition of variation: Principal component analysis (PCA).
- Recovery of $independent$ sources: Independent component analysis (ICA)
- Other methods: nonlinear PCA, kernel PCA...etc

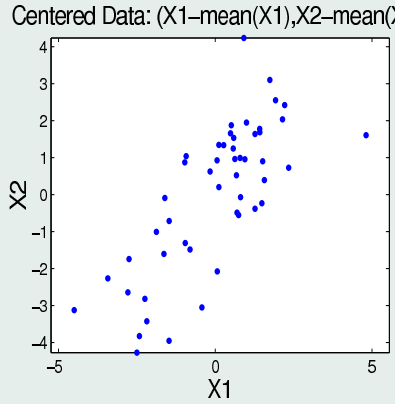
1. PCA

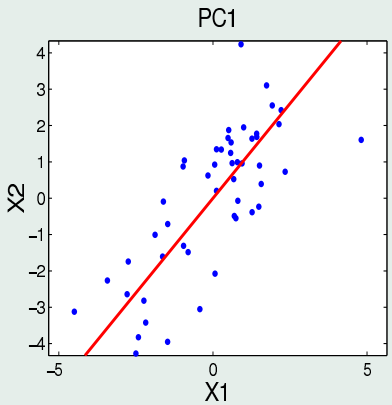
other names

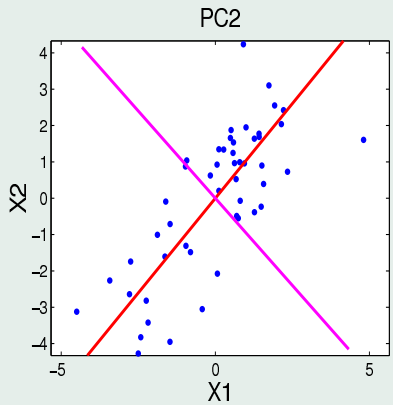
- Factor Analysis
- Karhunen-Loeve expansion
- Proper Orthogonal Decomposition (POD)

PCA for $d = 2$, $X = (X_1, X_2)$

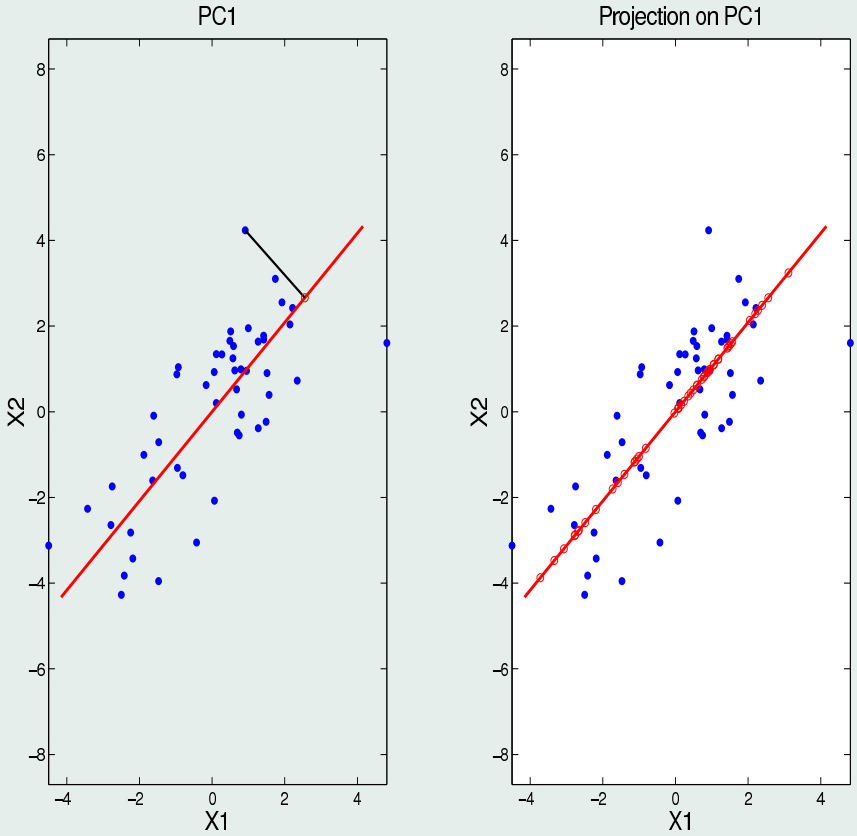




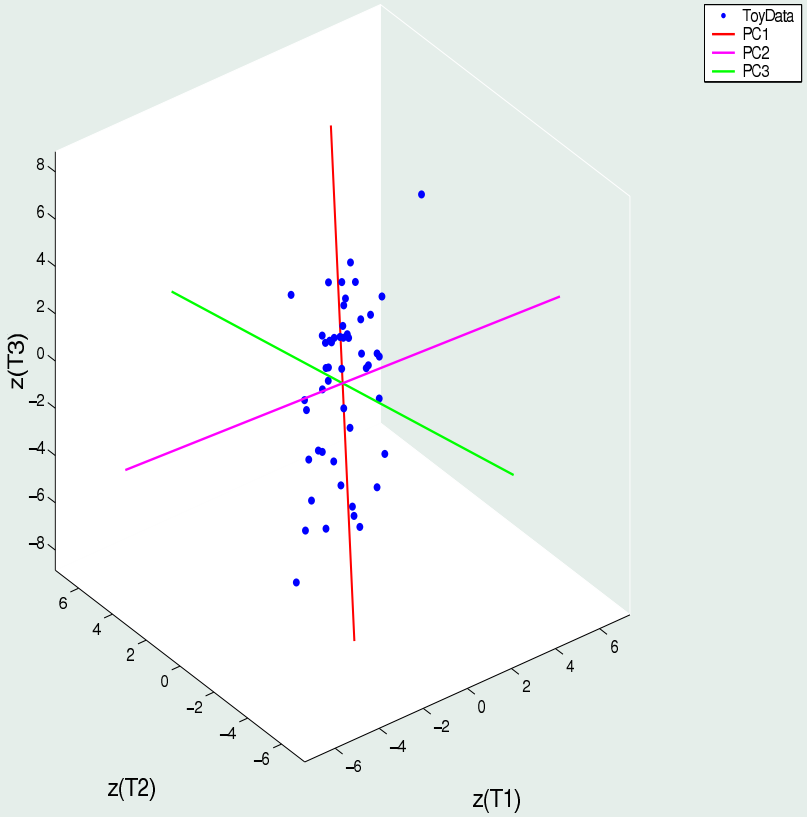




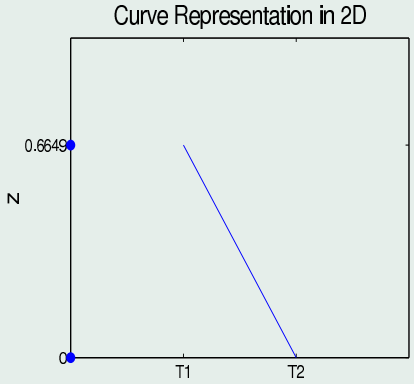
PC1, direction of greatest variability

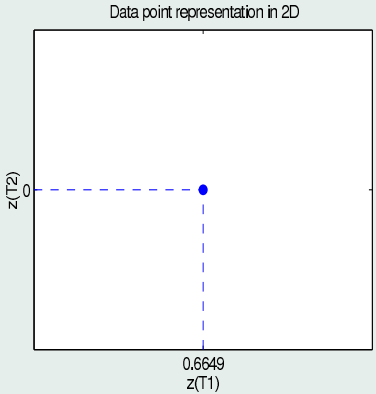


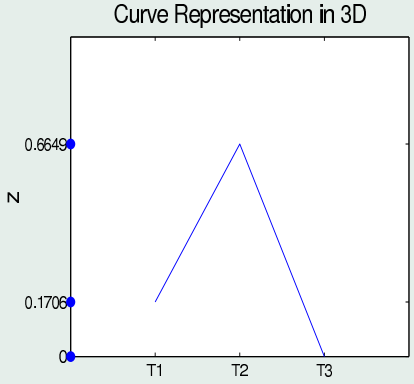
PCA for $d = 3$, $X = (X_1, X_2, X_3)$. movie

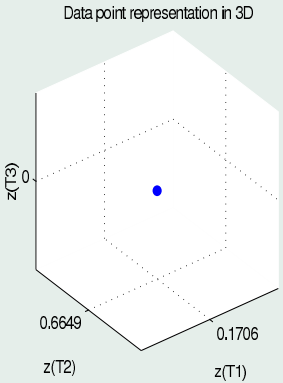


Curve representation vs point cloud representation

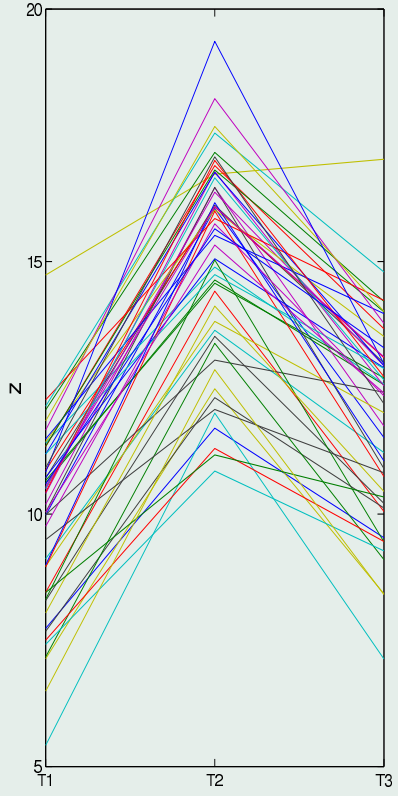


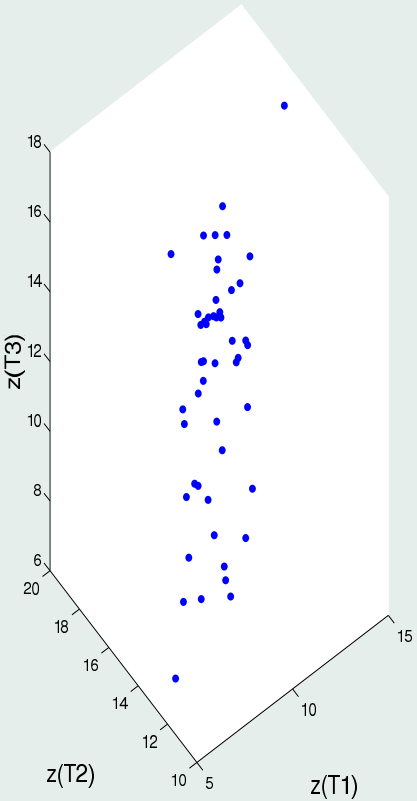




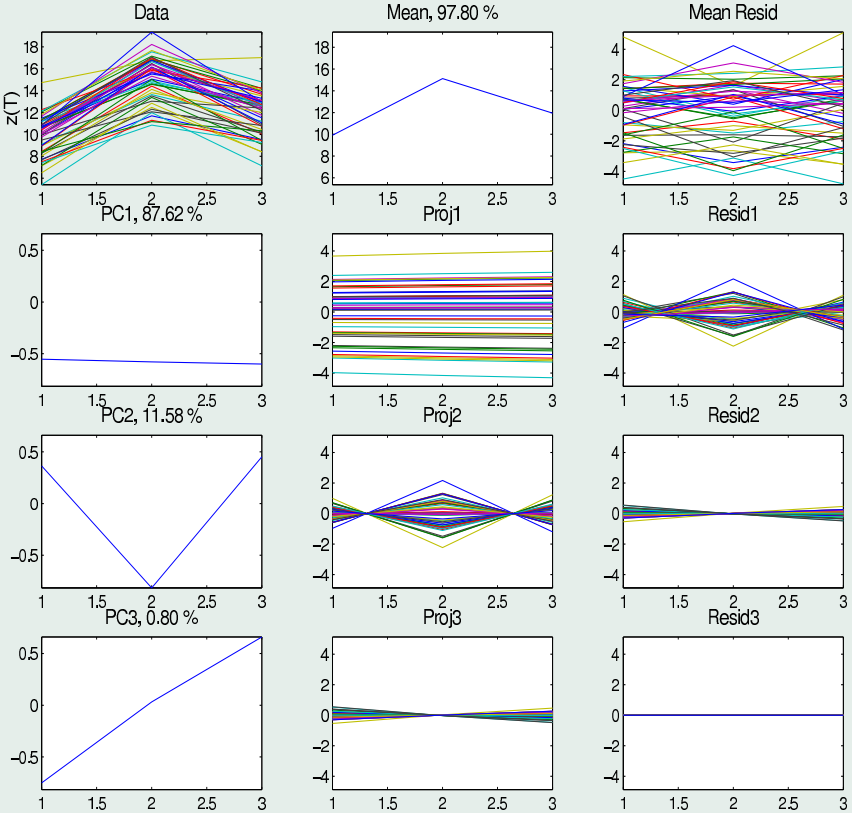


Variation in Curves vs Variation in point cloud

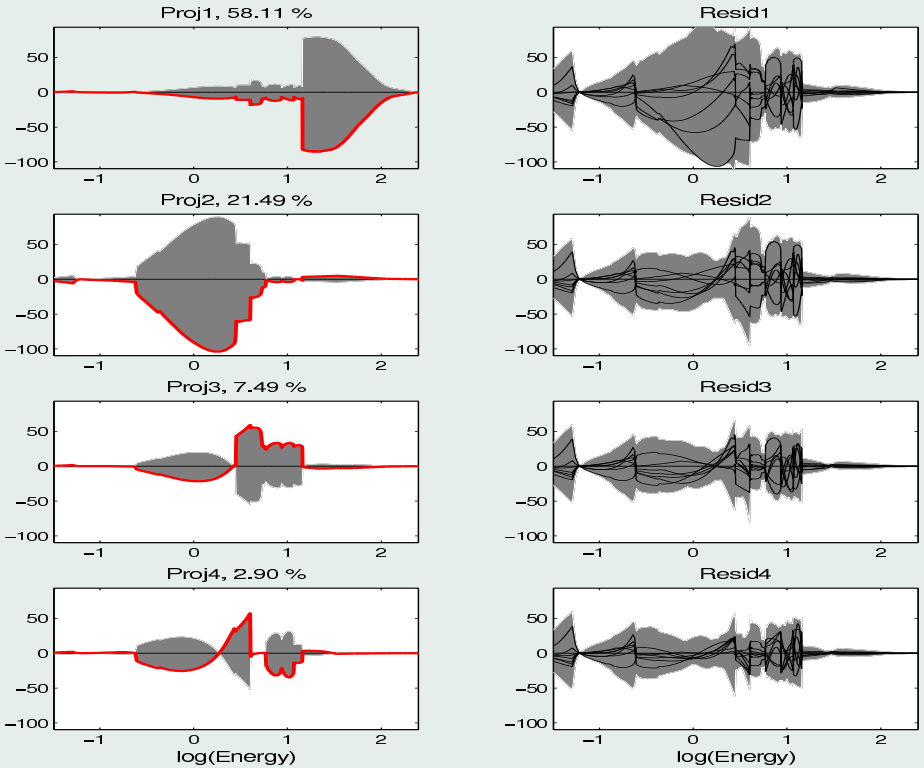




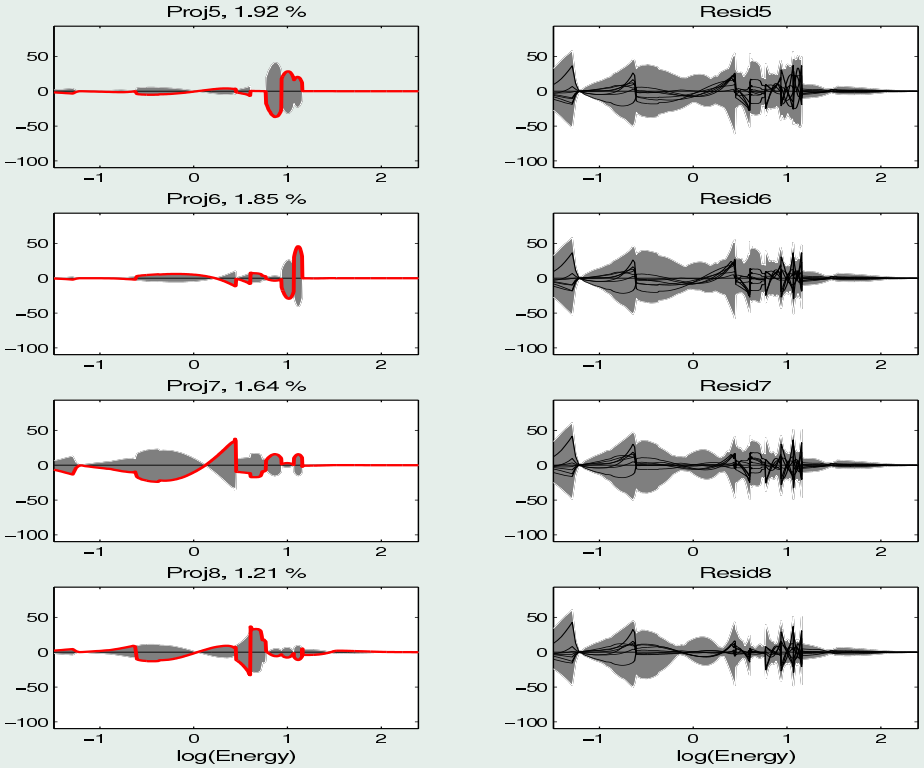
PCA, Curve representation



Results on data of interest



Results on data of interest (contd)



Results on data of interest (contd)

- We reduced the dimensionality from 1078 to 8 significant directions
- The first 4 PC's explain about 90% of the total variation in the data.
- PC1: positive correlation at high energy.
- PC2: positive correlation at medium energy
- PC3 negative correlation btw medium low energy and medium high energy.

PCA, orthog'l directions of greatest variability

- PC1 max. the variance, its direction $P_1 = Argmax_{(\alpha \in \mathbb{R}^d, \alpha'\alpha=1)}(\alpha'\Sigma\alpha)$.
- PC2 max. the variance in the space $\perp$ PC1, its direction $P_2 = Argmax_{(\alpha \in \mathbb{R}^d, ||\alpha||=1, \alpha' P_1=0)}(\alpha'\Sigma\alpha)$.
- Repeat iteration to find $P_3, \ldots, P_d$

PCA, eigenvectors of var-cov matrix Σ
Let P be the matrix of eigenvectors of Σ, then

$$\Sigma = P\Lambda P'$$

where $\Lambda = diag(\lambda_1, \ldots, \lambda_d)$, λ_i is the eigenvalue associated with P_i and $\lambda_i \geq 0 \forall i$.

Deriving PCs from Σ

- Previous result: $P'\Sigma P = \Lambda = diag(\lambda_1, \ldots, \lambda_d)$
- If order Λ such as $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_d$, then P' is the matrix of principal directions i.e $P' = [P1, P2, \ldots, P3]$. So $\forall i, Pi$ is an eigenvector of Σ associated to the eigenvalue λ_i.

Quantifying the variation

- $RSS = \frac{Var(Pi'X)}{||Var(X)||}$ quantifies the variation along PCi
- Equivalently $RSS = \frac{\lambda_i}{\sum_{i=1}^{n} \lambda_i}$

Change of variables after dimensionality reduction,

- Let P'_{sig} be the $(d$ x $d')$ matrix of significant PCs associated to Λ_{sig}
- Note that, if $X \sim N(\mu, \Sigma)$, then $Z = \Lambda_{sig}^{-0.5} P_{sig} X$ is $N(0, Id_{d'})$.

2. ICA

For non-gaussian data. ICA of the random vector $\mathbf{x}$ consists of finding a linear transform $\mathbf{s} = \mathbf{W}\mathbf{x}$ so that the components s_i are as independent as possible, in the sense of maximizing some function $F(s_1, ..., s_{d'})$ that measures independence.

Main assumptions:

- All independent components s_i, with the possible exception of one component, must be non-Gaussian.
- Number of observed linear mixtures d must be at least as large as the number of independent components,d', i.e. $d > d'$
- the matrix W must be of full column rank.

Some examples of functions to optimize

- Likelihood function:

$$L = \sum_t \sum_i log(f_i(w_i' x(t))) + T \ln |\det W|$$

where f_i is density of s_i

- Mutual information :

$$I(s_1, \ldots, s_m) = \sum_i H(s_i) - H(s)$$

- Kullback-Leibler divergence

$$\delta(f, \tilde{f}) = \int f(s) log \frac{f(s)}{\tilde{f}(s)} ds$$

where $f(s)$ is the joint density of sources s, and $\tilde{f}(s)$ is the product of the marginals, i.e. $\tilde{f}(s) = \prod f_i(s_i)$.