Regularized PCA to denoise and visualise data

Marie Verbanck – **Julie Josse** – François Husson

Laboratoire de statistique, Agrocampus Ouest, Rennes, France

CNAM, Paris, 16 janvier 2013

Outline

Geometrical point of view

Maximize the variance of the projected points
Minimize the reconstruction error

$\Rightarrow$ Approximation of $\mathbf{X}$ of low rank ($S < p$)

$$\|\mathbf{X}_{n\times p} - \hat{\mathbf{X}}_{n\times p}\|^2 \qquad \text{SVD: } \hat{\mathbf{X}}^{PCA} = \mathbf{U}_{n\times S}\Lambda^{\frac{1}{2}}_{S\times S}\mathbf{V}'_{p\times S}$$

$\mathbf{F} = \mathbf{U}\Lambda^{\frac{1}{2}}$ principal components
$\mathbf{V}$ principal axes

Alternating least squares algorithm: $\|\mathbf{X} - \mathbf{F}\mathbf{V}'\|^2$

$$\mathbf{F} = \mathbf{X}\mathbf{V}(\mathbf{V}'\mathbf{V})^{-1}$$
$$\mathbf{V} = \mathbf{X}'\mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}$$

Model in PCA

$$x_{ij} = \text{signal} + \text{noise}$$

- Random effect model: Factor Analysis, Probabilistic PCA (Lawley, 1953, Roweis, 1998, Tipping & Bishop, 1999)
  $\Rightarrow$ individuals are exchangeable
  $\Rightarrow$ relationship between variables

- Fixed effect model (Lawley, 1941, Caussinus 1986)
  $\Rightarrow$ individuals are not i.i.d
  $\Rightarrow$ study the individuals and the variables
  - sensory analysis (products - sensory descriptors)
  - plant breeding (genotypes - environments)

Fixed effect model

$$\mathbf{X}_{n\times p} = \tilde{\mathbf{X}}_{n\times p} + \varepsilon_{n\times p}$$

$$x_{ij} = \sum_{s=1}^{S} \sqrt{d_s} q_{is} r_{js} + \varepsilon_{ij} \ , \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$$

- Randomness due to the noise
- Individuals have different expectations: individual parameters
- Inferential framework: when $n \nearrow$ the number of parameters $\nearrow$
  $\Rightarrow$ Asymptotic: $\sigma \to 0$
- Maximum likelihood estimates $(\hat{\theta})$ = least square estimates
  EM algorithm = ALS algorithm
  $\Rightarrow$ Not necessarily the best in MSE

$\Rightarrow$ Shrinkage estimates: minimizes $\mathbb{E}(\phi\hat{\theta} - \theta)^2$

Outline

Minimizing the MSE

$$x_{ij} = \tilde{x}_{ij} + \varepsilon_{ij} = \sum_{s=1}^{S} \tilde{x}_{ij}^{s} + \varepsilon_{ij}$$

$$\text{MSE} = \mathbb{E}\left(\sum_{i,j}\left(\sum_{s=1}^{\min(n-1,p)} \phi_s \hat{x}_{ij}^{(s)} - \tilde{x}_{ij}\right)^2\right) \text{ with } \hat{x}_{ij}^{(s)} = \sqrt{\lambda_s} u_{is} v_{js}$$

$$\text{MSE} = \mathbb{E}\left(\sum_{i,j}\left(\sum_{s=1}^{S} \phi_s \hat{x}_{ij}^{(s)} - \tilde{x}_{ij}^{(s)}\right)^2 + \sum_{s=S+1}^{\min(n-1,p)}\left(\phi_s \hat{x}_{ij}^{(s)} - \tilde{x}_{ij}^{(s)}\right)^2\right)$$

For all $s \geq S+1$, $\tilde{x}_{ij}^{(s)} = 0$, therefore $\phi_{S+1} = \ldots = \phi_{\min(n-1,p)} = 0$.

Minimizing the MSE

$$\text{MSE} = \mathbb{E}\left(\sum_{i,j}\left(\sum_{s=1}^{S}\phi_s \hat{x}_{ij}^{(s)} - \tilde{x}_{ij}\right)^2\right) \text{ with } \hat{x}_{ij}^{(s)} = \sqrt{\lambda_s} u_{is} v_{js}$$

$$\phi_s = \frac{\sum_{i,j}\mathbb{E}(\hat{x}_{ij}^{(s)})\tilde{x}_{ij}}{\sum_{i,j}\mathbb{E}(\hat{x}_{ij}^{(s)2})} = \frac{\sum_{i,j}\mathbb{E}(\hat{x}_{ij}^{(s)})\tilde{x}_{ij}}{\sum_{i,j}\left(V(\hat{x}_{ij}^{(s)}) + \left(\mathbb{E}(\hat{x}_{ij}^{(s)})\right)^2\right)}$$

When $\sigma \to 0$: $\mathbb{E}(\hat{x}_{ij}^{(s)}) = \tilde{x}_{ij}^{(s)}$ (unbiased); $V(\hat{x}_{ij}^{(s)}) = \frac{\sigma^2}{\min(n-1;p)}$
(Pazman & Denis, 1996; Denis & Gower, 1999)

$$\phi_s = \frac{\sum_{i,j}\tilde{x}_{ij}^{(s)^2}}{\sum_{i,j}\frac{\sigma^2}{\min(n-1;p)} + \sum_{i,j}\tilde{x}_{ij}^{(s)^2}}$$

Minimizing the MSE

$$x_{ij} = \tilde{x}_{ij} + \varepsilon_{ij} = \sum_{s=1}^{S} \sqrt{d_s} q_{is} r_{js} + \varepsilon_{ij} \ , \ \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$$

$$\text{MSE} = \mathbb{E}\left(\sum_{i,j} \left(\sum_{s=1}^{S} \phi_s \hat{x}_{ij}^{(s)} - \tilde{x}_{ij} \right)^2 \right) \text{ with } \hat{x}_{ij}^{(s)} = \sqrt{\lambda_s} u_{is} v_{js}$$

$$\phi_s = \frac{\sum_{i,j} \tilde{x}_{ij}^{(s)^2}}{\sum_{i,j} \frac{\sigma^2}{\min(n-1;p)} + \sum_{i,j} \tilde{x}_{ij}^{(s)^2}} = \frac{d_s}{\frac{np\sigma^2}{\min(n-1;p)} + d_s}$$

$$= \frac{\text{variance signal}}{\text{variance total}}$$

$$\hat{\phi}_s = \frac{\lambda_s - \hat{\sigma}^2}{\lambda_s} \text{ for } s = 1, ..., S$$

Regularized PCA

$$\hat{x}_{ij}^{\text{PCA}} = \sum_{s=1}^{S} \sqrt{\lambda_s} u_{is} v_{js}$$

$$\hat{x}_{ij}^{\text{rPCA}} = \sum_{s=1}^{S} \left(\frac{\lambda_s - \hat{\sigma}^2}{\lambda_s} \right) \sqrt{\lambda_s} u_{is} v_{js} = \sum_{s=1}^{S} \left(\sqrt{\lambda_s} - \frac{\hat{\sigma}^2}{\sqrt{\lambda_s}} \right) u_{is} v_{js}$$

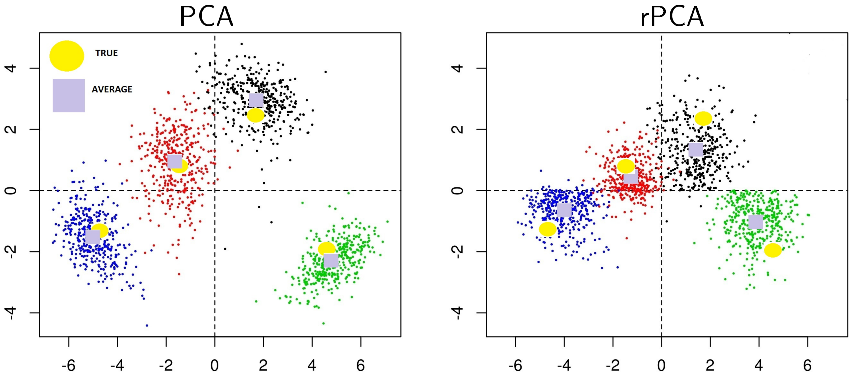
Compromise between hard and soft thresholding
σ^2 small $\rightarrow$ regularized PCA $\approx$ PCA
σ^2 large $\rightarrow$ individuals coordinates toward the center of gravity

$$\hat{\sigma}^2 = \frac{RSS}{\text{ddl}} = \frac{n \sum_{s=S+1}^{q} \lambda_s}{np - p - nS - pS + S^2 + S} \qquad (\mathbf{X}_{n \times p}; \mathbf{U}_{n \times S}; \mathbf{V}_{p \times S})$$

(only asympt unbiased)

Illustration of the regularization

$$\mathbf{X}^{sim}_{4\times10} = \tilde{\mathbf{X}}_{4\times10} + \varepsilon^{sim}_{4\times10} \qquad sim = 1, ..., 1000$$



$\Rightarrow$ rPCA more biased less variable

Outline

1. PCA
2. Regularized PCA
    - MSE point of view
    - Bayesian point of view
3. Results

Outline

1. PCA
2. Regularized PCA
   - MSE point of view
   - Bayesian point of view

$$
\begin{aligned}
&Y = \mathbf{X}\beta + \varepsilon,\ \varepsilon \sim \mathcal{N}(0, \sigma^2) \\
&\beta \sim \mathcal{N}(0,\ \tau^2) \\
&\hat{\beta}^{\text{MAP}} = (\mathbf{X}'\mathbf{X} + \kappa)^{-1}\mathbf{X}'Y, \text{ with } \kappa = \sigma^2/\tau^2
\end{aligned}
$$

3. Results

Probabilistic PCA (*Roweis, 1998, Tipping & Bishop, 1999*)

⇒ A specific case of an exploratory factor analysis model

$$\mathbf{x}_{ij} = (\mathbf{Z}_{n\times S}\mathbf{B}'_{p\times S})_{ij} + \varepsilon_{ij},\ \mathbf{z}_i \sim \mathcal{N}(0,\mathbb{I}_S),\ \varepsilon_i \sim \mathcal{N}(0,\sigma^2\mathbb{I}_p)$$

- Conditional independence : $x_i|z_i \sim \mathcal{N}(\mathbf{B}z_i,\ \mathbb{I}_p)$
- Distribution for the observations:
  $\mathbf{x}_i \sim \mathcal{N}(0,\Sigma)$ with $\Sigma = \mathbf{B}\mathbf{B}' + \sigma^2\mathbb{I}_p$
- Explicit solution:
  - $\hat{\sigma}^2 = \frac{1}{p-S}\sum_{s=S+1}^{p}\lambda_s$
  - $\hat{\mathbf{B}} = \mathbf{V}(\Lambda - \sigma^2\mathbb{I}_S)^{\frac{1}{2}}$

Probabilistic PCA

$$\mathbf{x}_{ij} = (\mathbf{Z}_{n\times S}\mathbf{B}'_{p\times S})_{ij} + \varepsilon_{ij},\ \mathbf{z}_i \sim \mathcal{N}(0,\mathbb{I}_S),\ \varepsilon_i \sim \mathcal{N}(0,\sigma^2\mathbb{I}_p)$$

$\Rightarrow$ Random effect model: no individual parameters
$\Rightarrow$ BLUP $\mathbb{E}(\mathbf{z}_i|\mathbf{x}_i)$: $\hat{\mathbf{Z}} = \mathbf{X}\hat{\mathbf{B}}(\hat{\mathbf{B}}'\hat{\mathbf{B}} + \sigma^2\mathbb{I}_S)^{-1}$

$$\hat{\mathbf{X}}^{\mathrm{pPCA}} = \hat{\mathbf{Z}}\hat{\mathbf{B}}'$$

Using $\hat{\mathbf{B}} = \mathbf{V}(\Lambda - \sigma^2\mathbb{I}_S)^{\frac{1}{2}}$ and SVD $\mathbf{X} = \mathbf{U}\Lambda^{1/2}\mathbf{V}' \rightarrow \mathbf{X}\mathbf{V} = \Lambda^{1/2}\mathbf{U}$

$$\hat{\mathbf{X}}^{\mathrm{pPCA}} = \mathbf{U}(\Lambda - \sigma^2\mathbb{I}_S)\Lambda^{-\frac{1}{2}}\mathbf{V} \qquad \hat{x}_{ij}^{\mathrm{pPCA}} = \sum_{s=1}^{S}\left(\sqrt{\lambda_s} - \frac{\hat{\sigma}^2}{\sqrt{\lambda_s}}\right)u_{is}v_{js}$$

Bayesian treatment of the fixed effect model with an a priori on z_i:
$\Rightarrow$ Regularized PCA

Probabilistic PCA

$\Rightarrow$ EM algorithm in pPCA: two ridge regressions

$$\begin{aligned} \text{Step E:} \quad & \hat{\mathbf{Z}} = \mathbf{X}\hat{\mathbf{B}}(\hat{\mathbf{B}}'\hat{\mathbf{B}} + \hat{\sigma}^2 \mathbb{I}_S)^{-1} \\ \text{Step M:} \quad & \hat{\mathbf{B}} = \mathbf{X}'\hat{\mathbf{Z}}(\hat{\mathbf{Z}}'\hat{\mathbf{Z}} + \hat{\sigma}^2 \Lambda^{-1})^{-1} \end{aligned}$$

$\Rightarrow$ EM algorithm in PCA: two regressions $\|\mathbf{X} - \mathbf{F}\mathbf{V}'\|^2$

$$\begin{aligned} \mathbf{F} &= \mathbf{X}\mathbf{V}(\mathbf{V}'\mathbf{V})^{-1} \\ \mathbf{V} &= \mathbf{X}'\mathbf{F}(\mathbf{F}'\mathbf{F})^{-1} \end{aligned}$$

An empirical Bayesian approach

$$\text{Model: } x_{ij} = \sum_{s=1}^{S} \tilde{x}_{ij}^{(s)} + \varepsilon_{ij} \text{ , } \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$$

$$\text{Prior: } \tilde{x}_{ij}^{(s)} \sim \mathcal{N}(0, \tau_s^2)$$

$$\text{Posterior: } \mathbb{E}\left(\tilde{x}_{ij}^{(s)} | x_{ij}^{(s)}\right) = \Phi_s x_{ij}^{(s)} \text{ with } \Phi_s = \frac{\tau_s^2}{\tau_s^2 + \frac{1}{\min(n-1;p)}\sigma^2}$$

Maximizing the likelihood of $\left(x_{ij}^{(s)}\right) \sim \mathcal{N}(0, \tau_s^2 + \frac{1}{\min(n-1;p)}\sigma^2)$ as a function of τ_s^2 to obtain: $\hat{\tau_s}^2 = \left(\frac{1}{np}\lambda_s - \frac{1}{\min(n-1;p)}\hat{\sigma}^2\right)$

$$\hat{\Phi}_s = \frac{\lambda_s - \frac{np}{\min(n-1;p)}\hat{\sigma}^2}{\lambda_s}$$

Outline

Simulation design

The simulated data:

$$\mathbf{X} = \tilde{\mathbf{X}} + \varepsilon$$

Vary several parameters to construct $\tilde{\mathbf{X}}$:

- n/p: $100/20 = 5$, $50/50 = 1$ and $20/100 = 0.2$
- S underlying dimensions: 2, 4, 10
- the ratio d_1/d_2: 4, 1

ε_{ij} is drawn from a Gaussian distribution with a standard deviation chosen to obtain a signal-to-noise ratio (SNR): 4, 1, 0.8

Soft thresholding and SURE (Candès *et al.*, 2012)

$$\hat{x}_{ij}^{\text{Soft}} = \sum_{s=1}^{\min(n,p)} \left(\sqrt{\lambda_s} - \lambda\right)_+ u_{is} v_{js} \qquad \|\mathbf{X} - \tilde{\mathbf{X}}\|^2 + \lambda \|\tilde{\mathbf{X}}\|_*$$

Choosing the tunning parameter λ? $\text{MSE}(\lambda) = \mathbb{E}\|\tilde{\mathbf{X}} - \text{SVT}_\lambda(\mathbf{X})\|^2$

$\Rightarrow$ Stein's Unbiased Risk Estimate (SURE)

$$\text{SURE}(\text{SVT}_\lambda)(\mathbf{X}) = -np\sigma^2 + \sum_{i}^{\min(n,p)} \min(\lambda^2, \lambda_i) + 2\sigma^2 \text{div}(\text{SVT}_\lambda)(\mathbf{X})$$

$\Rightarrow$ λ is obtained minimizing SURE. Remark: $\hat{\sigma}^2$ is required

Simulation study

n/p	S	SNR	d_1/d_2	$MSE(\hat{\mathbf{X}}^{\mathrm{PCA}}, \tilde{\mathbf{X}})$	$MSE(\hat{\mathbf{X}}^{\mathrm{rPCA}}, \tilde{\mathbf{X}})$	$MSE(\hat{\mathbf{X}}^{\mathrm{SURE}}, \tilde{\mathbf{X}})$
100/20	4	4	4	8.25E-04	**8.24E-04**	1.42E-03
100/20	4	4	1	8.26E-04	**8.25E-04**	1.43E-03
100/20	4	1	4	2.60E-01	**1.96E-01**	2.43E-01
100/20	4	1	1	2.47E-01	**2.04E-01**	2.62E-01
100/20	4	0.8	4	7.41E-01	**4.27E-01**	4.36E-01
100/20	4	0.8	1	6.68E-01	**4.40E-01**	4.83E-01
50/50	4	4	4	5.48E-04	**5.48E-04**	1.04E-03
50/50	4	4	1	5.46E-04	**5.46E-04**	1.04E-03
50/50	4	1	4	1.75E-01	**1.53E-01**	2.00E-01
50/50	4	1	1	1.68E-01	**1.52E-01**	2.09E-01
50/50	4	0.8	4	5.07E-01	3.87E-01	**3.85E-01**
50/50	4	0.8	1	4.67E-01	**3.76E-01**	4.13E-01
20/100	4	4	4	8.28E-04	**8.27E-04**	1.41E-03
20/100	4	4	1	8.29E-04	**8.28E-04**	1.42E-03
20/100	4	1	4	2.55E-01	**1.97E-01**	2.45E-01
20/100	4	1	1	2.48E-01	**2.04E-01**	2.60E-01
20/100	4	0.8	4	7.13E-01	**4.15E-01**	4.37E-01
20/100	4	0.8	1	6.66E-01	**4.34E-01**	4.78E-01

MSE in favor of rPCA

Simulation study

n/p	S	SNR	d_1/d_2	$MSE(\hat{\mathbf{X}}^{\text{PCA}}, \tilde{\mathbf{X}})$	$MSE(\hat{\mathbf{X}}^{\text{rPCA}}, \tilde{\mathbf{X}})$	$MSE(\hat{\mathbf{X}}^{\text{SURE}}, \tilde{\mathbf{X}})$
100/20	4	4	4	8.25E-04	**8.24E-04**	1.42E-03
100/20	4	4	1	8.26E-04	**8.25E-04**	1.43E-03
100/20	4	1	4	2.60E-01	**1.96E-01**	2.43E-01
100/20	4	1	1	2.47E-01	**2.04E-01**	2.62E-01
100/20	4	0.8	4	7.41E-01	**4.27E-01**	4.36E-01
100/20	4	0.8	1	6.68E-01	**4.40E-01**	4.83E-01
50/50	4	4	4	5.48E-04	**5.48E-04**	1.04E-03
50/50	4	4	1	5.46E-04	**5.46E-04**	1.04E-03
50/50	4	1	4	1.75E-01	**1.53E-01**	2.00E-01
50/50	4	1	1	1.68E-01	**1.52E-01**	2.09E-01
50/50	4	0.8	4	5.07E-01	3.87E-01	**3.85E-01**
50/50	4	0.8	1	4.67E-01	**3.76E-01**	4.13E-01
20/100	4	4	4	8.28E-04	**8.27E-04**	1.41E-03
20/100	4	4	1	8.29E-04	**8.28E-04**	1.42E-03
20/100	4	1	4	2.55E-01	**1.97E-01**	2.45E-01
20/100	4	1	1	2.48E-01	**2.04E-01**	2.60E-01
20/100	4	0.8	4	7.13E-01	**4.15E-01**	4.37E-01
20/100	4	0.8	1	6.66E-01	**4.34E-01**	4.78E-01

rPCA $\geq$ PCA when SNR is small - rPCA $\approx$ PCA when SNR is high

Simulation study

n/p	S	SNR	d_1/d_2	$MSE(\hat{\mathbf{X}}^{\text{PCA}}, \tilde{\mathbf{X}})$	$MSE(\hat{\mathbf{X}}^{\text{rPCA}}, \tilde{\mathbf{X}})$	$MSE(\hat{\mathbf{X}}^{\text{SURE}}, \tilde{\mathbf{X}})$
100/20	4	4	4	8.25E-04	**8.24E-04**	1.42E-03
100/20	4	4	1	8.26E-04	**8.25E-04**	1.43E-03
100/20	4	1	4	2.60E-01	**1.96E-01**	2.43E-01
100/20	4	1	1	2.47E-01	**2.04E-01**	2.62E-01
100/20	4	0.8	4	7.41E-01	**4.27E-01**	4.36E-01
100/20	4	0.8	1	6.68E-01	**4.40E-01**	4.83E-01
50/50	4	4	4	5.48E-04	**5.48E-04**	1.04E-03
50/50	4	4	1	5.46E-04	**5.46E-04**	1.04E-03
50/50	4	1	4	1.75E-01	**1.53E-01**	2.00E-01
50/50	4	1	1	1.68E-01	**1.52E-01**	2.09E-01
50/50	4	0.8	4	5.07E-01	3.87E-01	**3.85E-01**
50/50	4	0.8	1	4.67E-01	**3.76E-01**	4.13E-01
20/100	4	4	4	8.28E-04	**8.27E-04**	1.41E-03
20/100	4	4	1	8.29E-04	**8.28E-04**	1.42E-03
20/100	4	1	4	2.55E-01	**1.97E-01**	2.45E-01
20/100	4	1	1	2.48E-01	**2.04E-01**	2.60E-01
20/100	4	0.8	4	7.13E-01	**4.15E-01**	4.37E-01
20/100	4	0.8	1	6.66E-01	**4.34E-01**	4.78E-01

rPCA $\geq$ rPCA when $d_1 \geq d_2$

Simulation study

n/p	S	SNR	d_1/d_2	$MSE(\hat{\mathbf{X}}^{\text{PCA}}, \tilde{\mathbf{X}})$	$MSE(\hat{\mathbf{X}}^{\text{rPCA}}, \tilde{\mathbf{X}})$	$MSE(\hat{\mathbf{X}}^{\text{SURE}}, \tilde{\mathbf{X}})$
100/20	4	4	4	8.25E-04	**8.24E-04**	1.42E-03
100/20	4	4	1	8.26E-04	**8.25E-04**	1.43E-03
100/20	4	1	4	2.60E-01	**1.96E-01**	2.43E-01
100/20	4	1	1	2.47E-01	**2.04E-01**	2.62E-01
100/20	4	0.8	4	7.41E-01	**4.27E-01**	4.36E-01
100/20	4	0.8	1	6.68E-01	**4.40E-01**	4.83E-01
50/50	4	4	4	5.48E-04	**5.48E-04**	1.04E-03
50/50	4	4	1	5.46E-04	**5.46E-04**	1.04E-03
50/50	4	1	4	1.75E-01	**1.53E-01**	2.00E-01
50/50	4	1	1	1.68E-01	**1.52E-01**	2.09E-01
50/50	4	0.8	4	5.07E-01	3.87E-01	**3.85E-01**
50/50	4	0.8	1	4.67E-01	**3.76E-01**	4.13E-01
20/100	4	4	4	8.28E-04	**8.27E-04**	1.41E-03
20/100	4	4	1	8.29E-04	**8.28E-04**	1.42E-03
20/100	4	1	4	2.55E-01	**1.97E-01**	2.45E-01
20/100	4	1	1	2.48E-01	**2.04E-01**	2.60E-01
20/100	4	0.8	4	7.13E-01	**4.15E-01**	4.37E-01
20/100	4	0.8	1	6.66E-01	**4.34E-01**	4.78E-01

Same order of errors for $n/p = 0.2$ or 5, smaller for $n/p = 1$!

Simulation study

n/p	S	SNR	d_1/d_2	$MSE(\hat{\mathbf{X}}^{\text{PCA}}, \tilde{\mathbf{X}})$	$MSE(\hat{\mathbf{X}}^{\text{rPCA}}, \tilde{\mathbf{X}})$	$MSE(\hat{\mathbf{X}}^{\text{SURE}}, \tilde{\mathbf{X}})$
100/20	4	4	4	8.25E-04	**8.24E-04**	1.42E-03
100/20	4	4	1	8.26E-04	**8.25E-04**	1.43E-03
100/20	4	1	4	2.60E-01	**1.96E-01**	2.43E-01
100/20	4	1	1	2.47E-01	**2.04E-01**	2.62E-01
100/20	4	0.8	4	7.41E-01	**4.27E-01**	4.36E-01
100/20	4	0.8	1	6.68E-01	**4.40E-01**	4.83E-01
50/50	4	4	4	5.48E-04	**5.48E-04**	1.04E-03
50/50	4	4	1	5.46E-04	**5.46E-04**	1.04E-03
50/50	4	1	4	1.75E-01	**1.53E-01**	2.00E-01
50/50	4	1	1	1.68E-01	**1.52E-01**	2.09E-01
50/50	4	0.8	4	5.07E-01	3.87E-01	**3.85E-01**
50/50	4	0.8	1	4.67E-01	**3.76E-01**	4.13E-01
20/100	4	4	4	8.28E-04	**8.27E-04**	1.41E-03
20/100	4	4	1	8.29E-04	**8.28E-04**	1.42E-03
20/100	4	1	4	2.55E-01	**1.97E-01**	2.45E-01
20/100	4	1	1	2.48E-01	**2.04E-01**	2.60E-01
20/100	4	0.8	4	7.13E-01	**4.15E-01**	4.37E-01
20/100	4	0.8	1	6.66E-01	**4.34E-01**	4.78E-01

SURE good when data are noisy

Simulation from Candes *et al.* (2012)

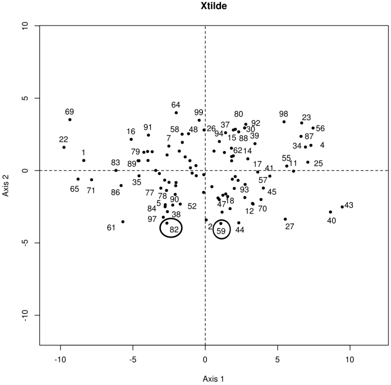
- Same model of simulation
- $n = 500$ $p = 200$, SNR (4, 2, 1, 0.5), S (10, 100)

S	SNR	$MSE(\hat{\mathbf{X}}^{\text{PCA}}, \tilde{\mathbf{X}})$	$MSE(\hat{\mathbf{X}}^{\text{rPCA}}, \tilde{\mathbf{X}})$	$MSE(\hat{\mathbf{X}}^{\text{SURE}}, \tilde{\mathbf{X}})$
10	4	4.31E-03	**4.29E-03**	8.74E-03
10	2	1.74E-02	**1.71E-02**	3.29E-02
10	1	7.16E-02	**6.75E-02**	1.16E-01
10	0.5	3.19E-01	**2.57E-01**	3.53E-01
100	4	3.79E-02	**3.69E-02**	4.50E-02
100	2	1.58E-01	**1.41E-01**	1.56E-01
100	1	7.29E-01	4.91E-01	**4.48E-01**
100	0.5	3.16E+00	1.48E+00	**8.52E-01**

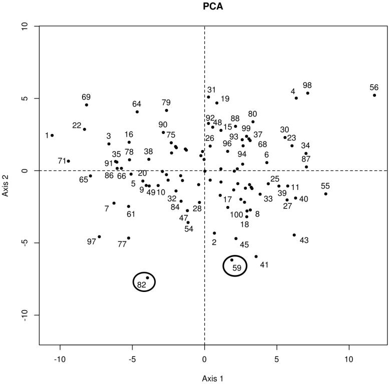
Individuals representation

100 individuals, 20 variables, 2 dimensions, SNR=0.8, $d_1/d_2 = 4$

True configuration



PCA

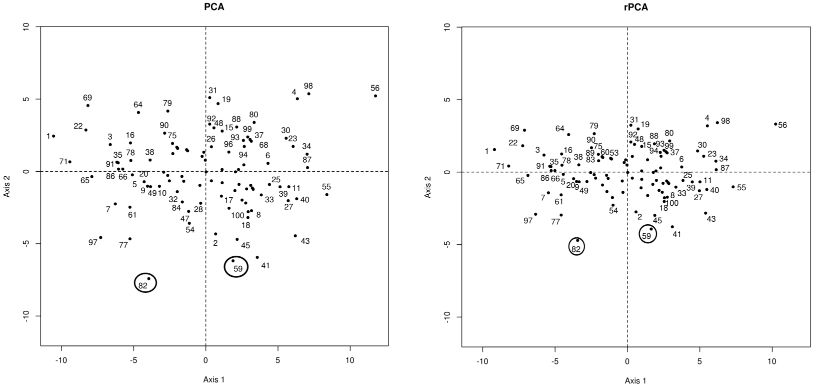


Individuals representation

100 individuals, 20 variables, 2 dimensions, SNR=0.8, $d_1/d_2 = 4$

PCA

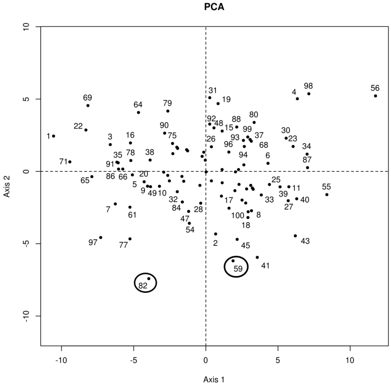
rPCA



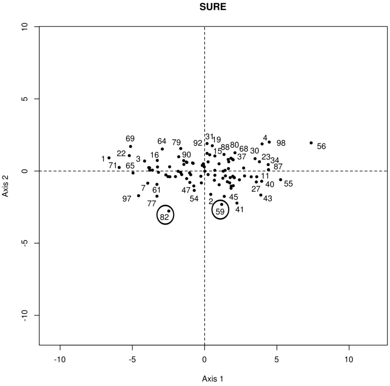
Individuals representation

100 individuals, 20 variables, 2 dimensions, SNR=0.8, $d_1/d_2 = 4$

PCA



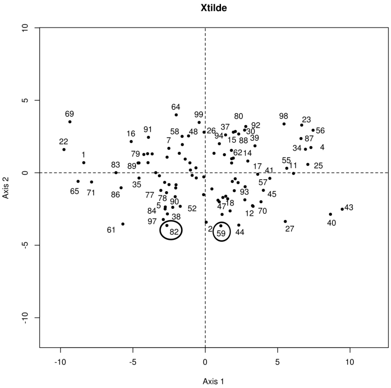
SURE



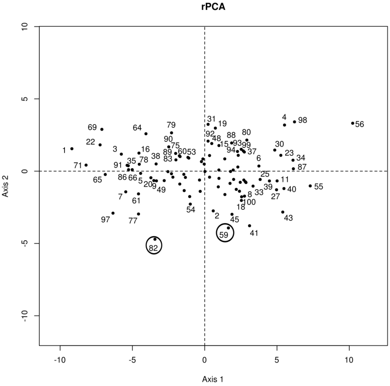
Individuals representation

100 individuals, 20 variables, 2 dimensions, SNR=0.8, $d_1/d_2 = 4$

True configuration



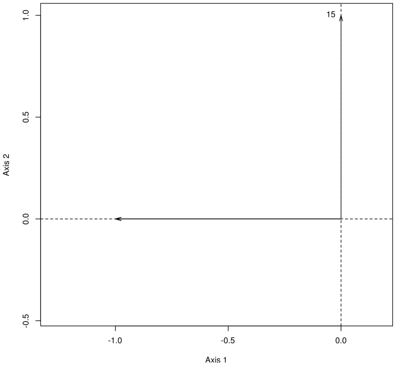
rPCA



$\Rightarrow$ Improve the recovery of the inner-product matrix $\tilde{\mathbf{X}}\tilde{\mathbf{X}}'$

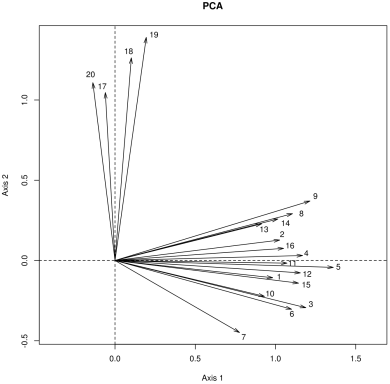
Variables representation

True configuration



$\mathrm{cor}(X_7,\ X_9) = 1$

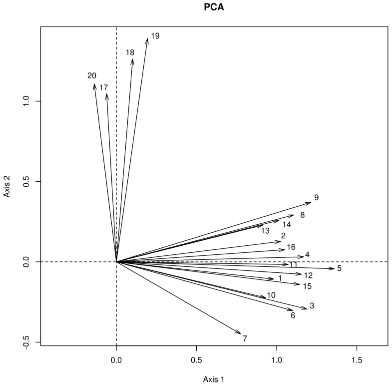
PCA



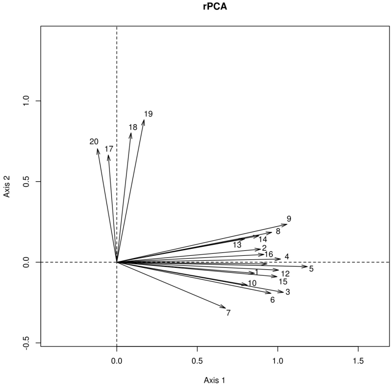
0.68

Variables representation

PCA



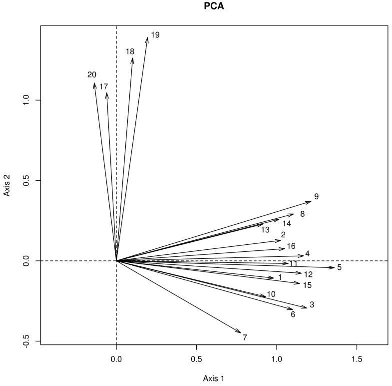
rPCA



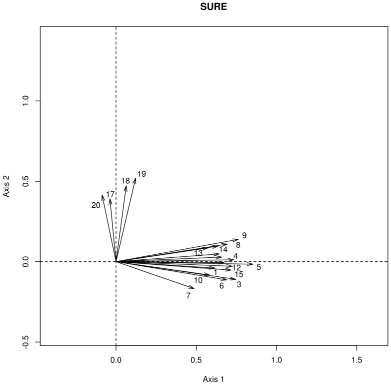
$\text{cor}(X_7, X_9) = 0.68$ 0.81

Variables representation

PCA



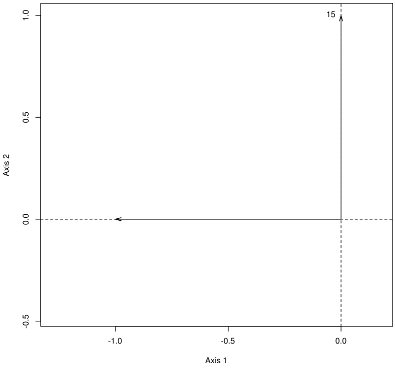
SURE



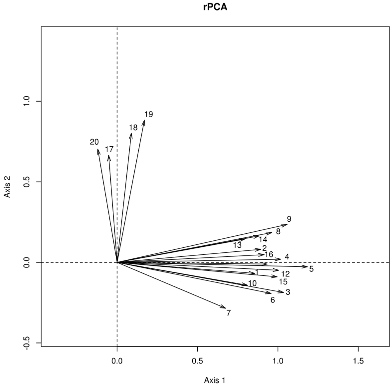
cor(X_7, X_9)= 0.68 0.82

Variables representation

True configuration



rPCA



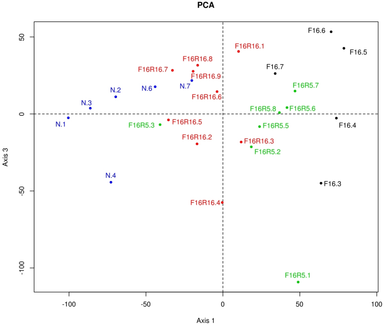
$\text{cor}(X_7, X_9) = 1$ 0.81

$\Rightarrow$ Improve the recovery of the covariance matrix $\tilde{\mathbf{X}}'\tilde{\mathbf{X}}$

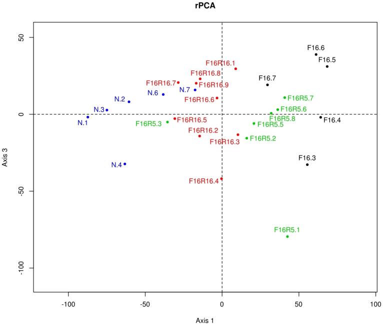
Real data set

- 27 chickens submitted to 4 nutritional statuses
- 12664 gene expressions

PCA



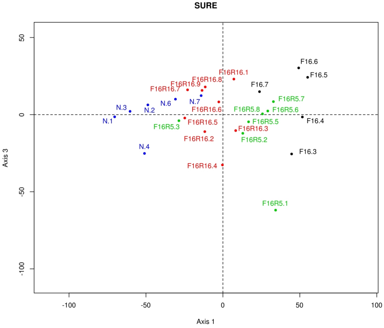
rPCA



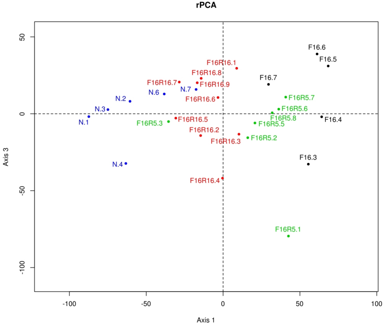
Real data set

- 27 chickens submitted to 4 nutritional statuses
- 12664 gene expressions

SURE



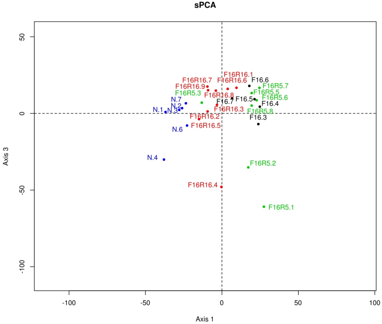
rPCA



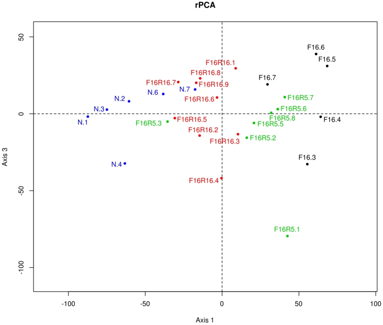
Real data set

- 27 chickens submitted to 4 nutritional statuses
- 12664 gene expressions

SPCA

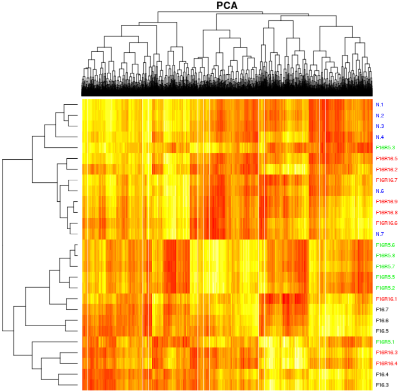


rPCA

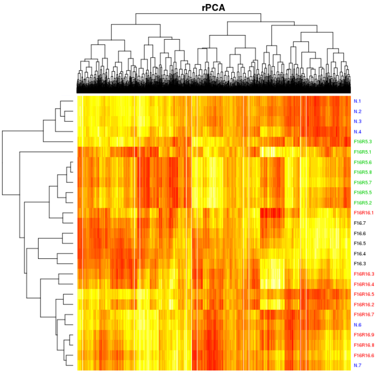


Real data set

PCA



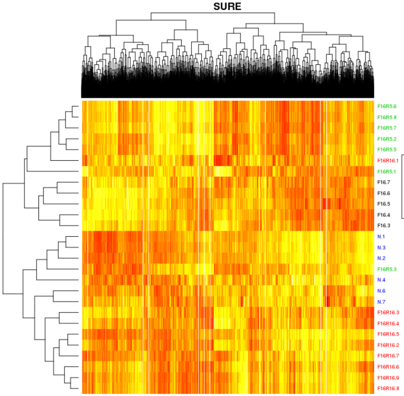
rPCA



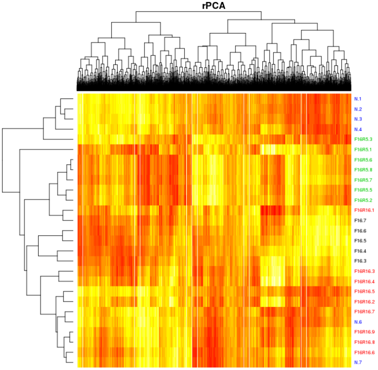
Bi-clustering from rPCA better find the 4 nutritional statuses!

Real data set

SURE



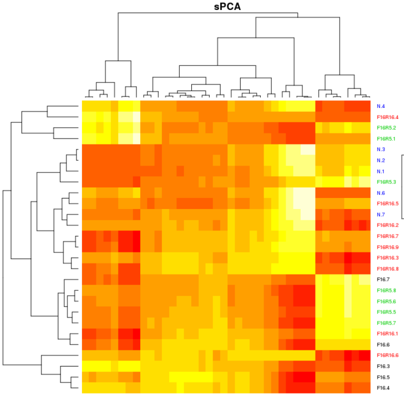
rPCA



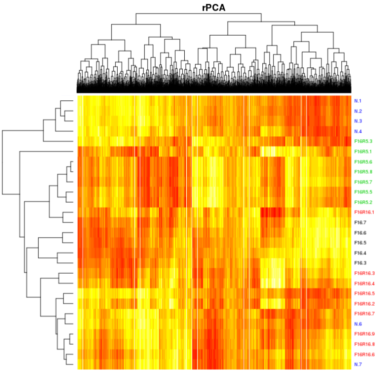
Bi-clustering from rPCA better find the 4 nutritional statuses!

Real data set

SPCA



rPCA



Bi-clustering from rPCA better find the 4 nutritional statuses!

Conclusion

$\Rightarrow$ Recover a low rank matrix from noisy observations

- singular values thresholding: a popular estimation strategy
- a regularized term a priori: $\left(\sqrt{\lambda_s} - \frac{\hat{\sigma}^2}{\sqrt{\lambda_s}}\right)$: Minimize MSE - Bayesian fixed effect model

$\Rightarrow$ Good results in term of recovery of the structure, the distances between individuals and the covariance matrix and in term of visualization

- Asymptotic results
- Tuning parameter: the number of dimensions $\rightarrow$ CV?

- Regularized MCA (Takane, 2006)