Singular Value Decomposition

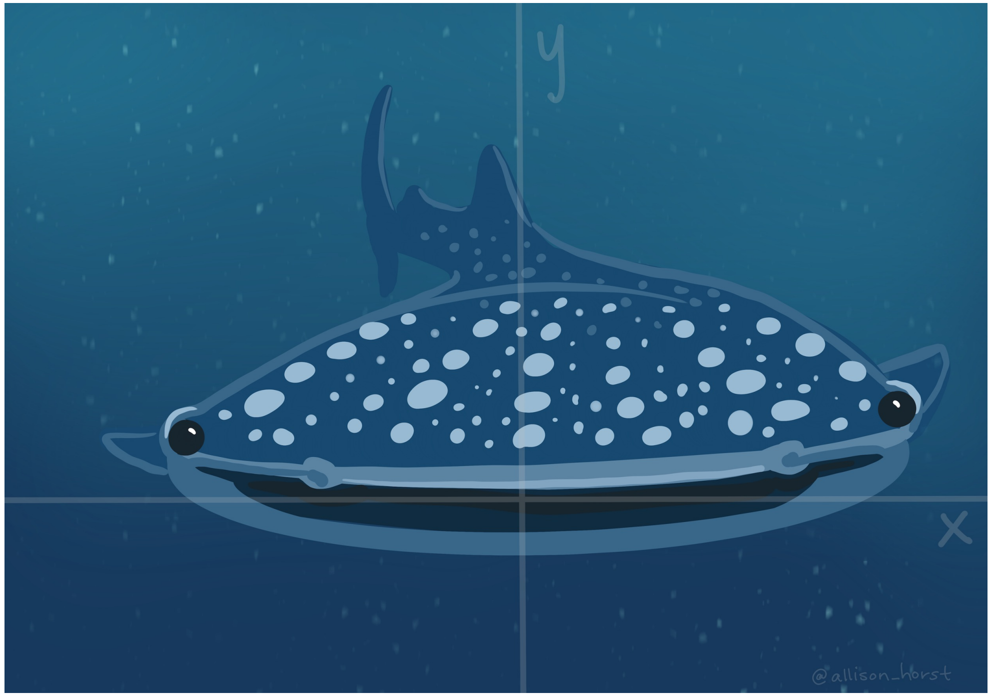
Vincent Guillemot

SOME MENTAL IMAGES

Potato Chips Analysis

Cut the yummiest French fries

Whale versus krill: this is you (credit: Allison Horst)



Eat the most krill (put on your 3D glasses)

Artwork by @allison_horst

Whale versus krill: this is your data (credit: Allison Horst)

Eat the most krill (put on your 3D glasses)

Artwork by @allison_horst

THE TRI-FORCE OF PCA

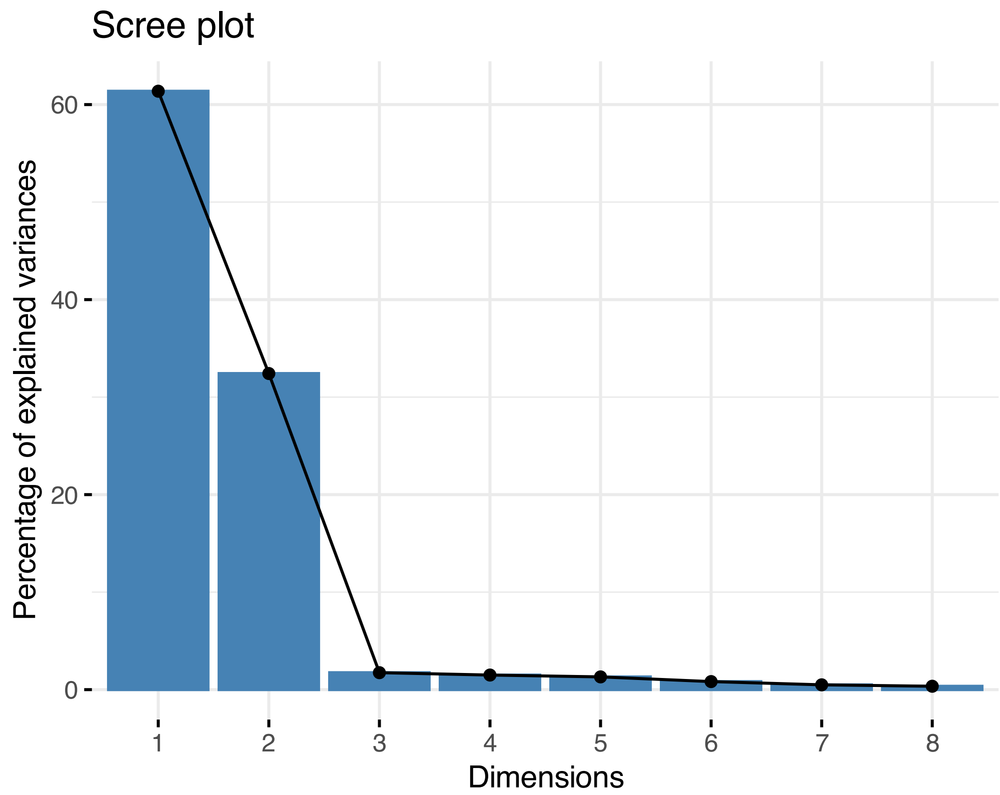
Beautiful illustration

Screeplot, versus individual map, versus circle of correlation. With the associated theoretical concepts: inertia (multivariate variance), distance between individuals, and angles between

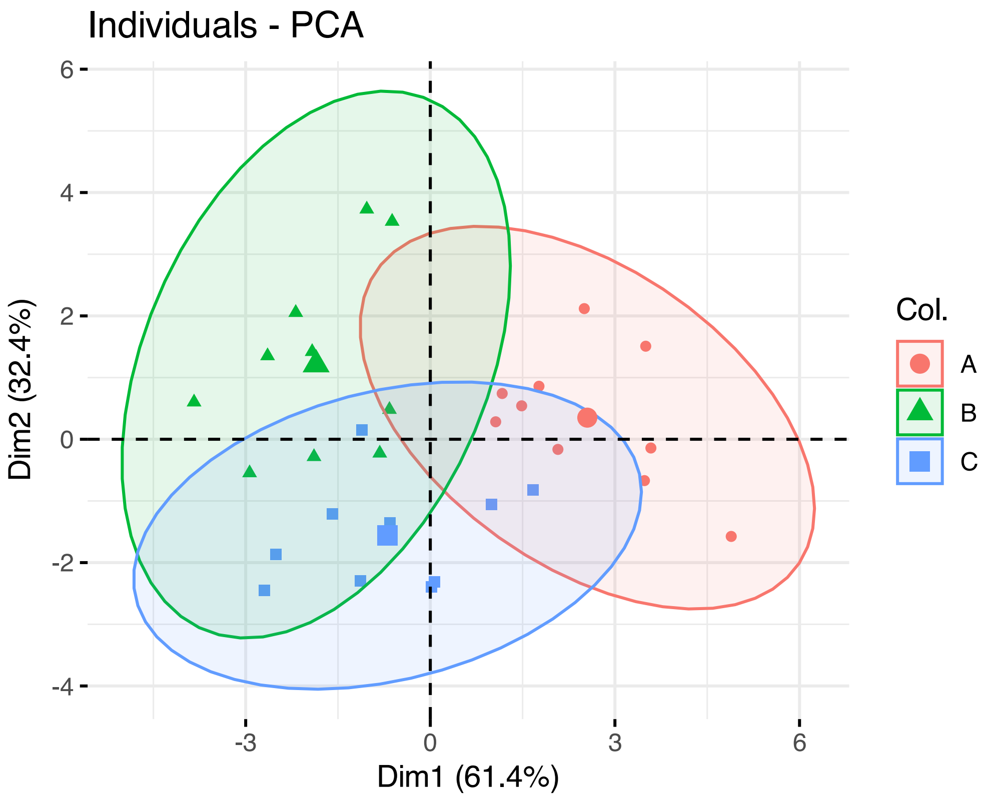
Example data

```
xtmp <-
readxl::read_excel("../data/simul.xlsx")
x <- as.matrix(xtmp[, -1])
rownames(x) <- xtmp$Ind
```

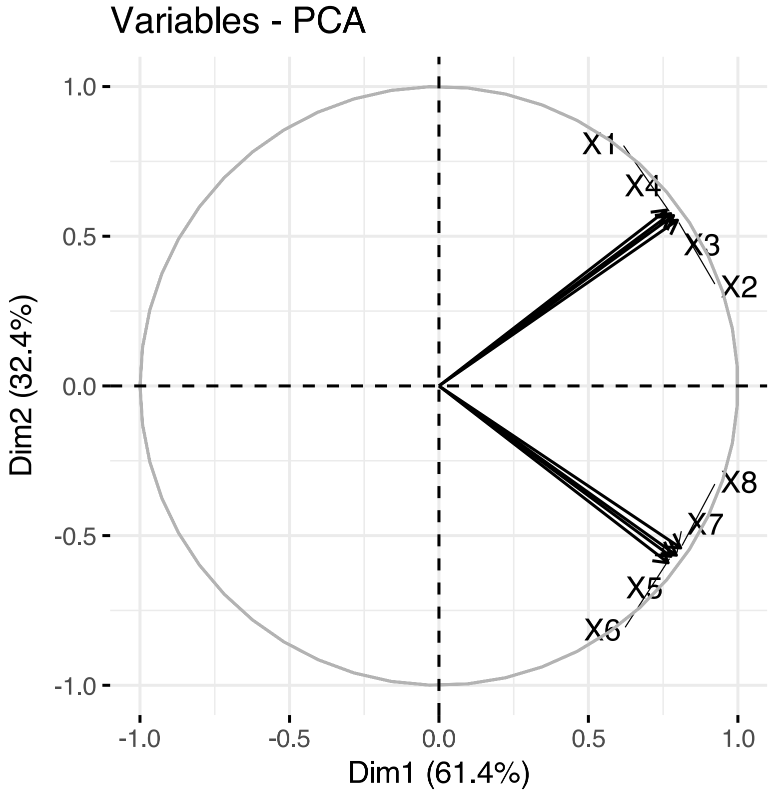
Example screeplot



Example individual map



Example circle of correlation



VOCABULARY

French versus English

“Aaaaah, mais ACP en fait c’est la PCA !”

(Anonymous student, after 6 hours of teaching PCA in French)

English	French
PCA = principal component analysis	ACP = analyse en composantes principales
SVD = singular value decomposition	SVD = décomposition en valeurs singulières
EVD = eigenvalue decomposition	décomposition en éléments propres
ICA = independent component analysis	ICA = analyse en composantes indépendantes
MDS = multidimensional scaling	MDS = multidimensional scaling

R vocabulary

Base methods:

- `eigen` for eigenvalue decomposition, `svd` for singular value decomposition,
- `prcomp` and `princomp` for PCA,
- `biplot`

Nice packages:

- `FactoMineR`: `PCA`, `MFA`, `CA`, `MCA` and associates. In earlier versions, the graphs were "crude"…
- `factoextra`: "helper" package to make beautiful plots, and much more!
- `ade4`: more than "one block" type of analyses. Made by ecologists so ⇒ PCOA, coinertia analysis, STATIS, etc.
- `ExPosition`: made for psychometricians (they like PLS)

And a few nice books and papers

books and papers

A LITTLE BIT OF MATH

Notations

(non-universal) Conventions: matrices and vectors are **bold**

- I = number of observations, J = number of variables (only quantitative)
- i for an individual observation, and j for a single variable
- $\mathbf{X}$ = data matrix, with n rows and p columns, sometimes already centered, and scaled, to make our life easy
- $\mathbf{X}_j$ = variable j, and jth column of $\mathbf{X}$
- $\mathbf{w}$ a set of weights

A little detour: matrix multiplication

Take a pen and paper, and do this multiplication:

$$\begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 2 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 2 & 3 & 4 \\ -1 & 0 & 0 & 1 \end{bmatrix}$$

[Cool video: 5 ways to see matrix multiplication]

PCA

"Find a linear combination of the columns of the data that would capture the most information."

In mathematical words, find

$$\mathbf{Xw} = w_1\mathbf{X}_1 + \cdots + w_p\mathbf{X}_p$$

that maximizes… wait a minute! What are the dimensions?

- $\mathbf{X}$: I rows and J columns,
- $\mathbf{w}$: J rows and 1 columns,
- $\mathbf{Xw}$: I rows and 1 column.

THE MATHEMATICAL TRANSLATION OF THE INTUITIONS BEHIND PCA

Most popular intuition of PCA: how does it translate?

"PCA creates a linear combination of variables that maximizes variance."

$$\arg\max_{\|\mathbf{w}\|_2^2=1} \mathrm{var}(\mathbf{X}\mathbf{w})$$

- Why $\| \mathbf{w} \|_2 = 1$?
- Dirty trick: $\mathrm{var}(\mathbf{X}\mathbf{w}) = \mathbf{w}^\top \mathbf{X}^\top \mathbf{X}\mathbf{w}$

Least “well-known” intuition of PCA: how does it translate?

“PCA creates a linear combination of variables that maximizes correlation.”

$$\underset{\mathbf{w}}{\operatorname{argmax}} \sum_{j=1}^{p} \operatorname{cor}\left(\mathbf{X}\mathbf{w}, \mathbf{X}_j\right)^2$$

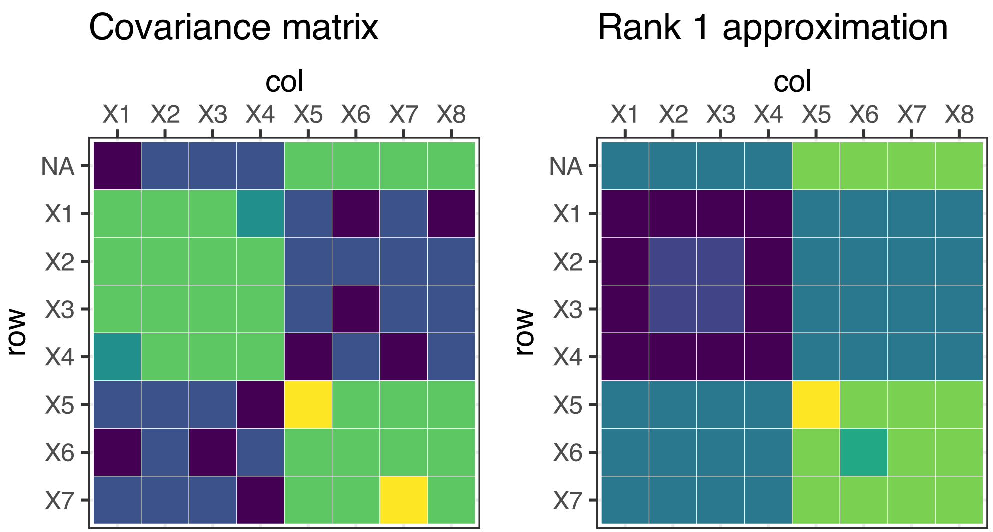
Second least “well-known” intuition of PCA: how does it translate?

“PCA creates the best lower rank approximation of the covariance matrix.”

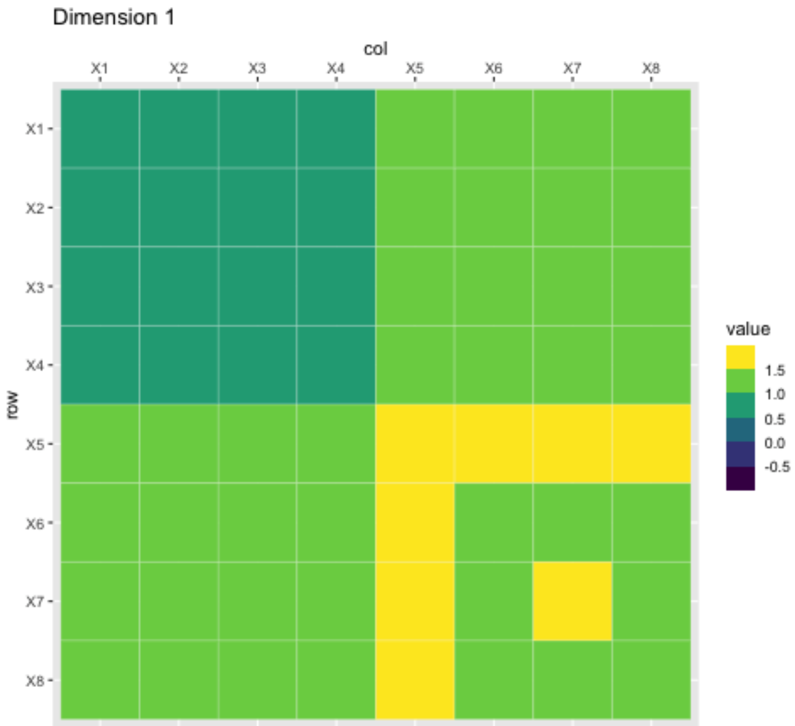
$$\arg\min_{\|w\|_2^2=1} \left\|\frac{1}{n}\mathbf{X}^\top\mathbf{X} - \lambda\mathbf{w}\mathbf{w}^\top\right\|_F^2$$

- $\frac{1}{n}\mathbf{X}^\top\mathbf{X}$
- λ: the [blank] of the covariance matrix
- $\mathbf{w}$: the [blank] of the covariance matrix

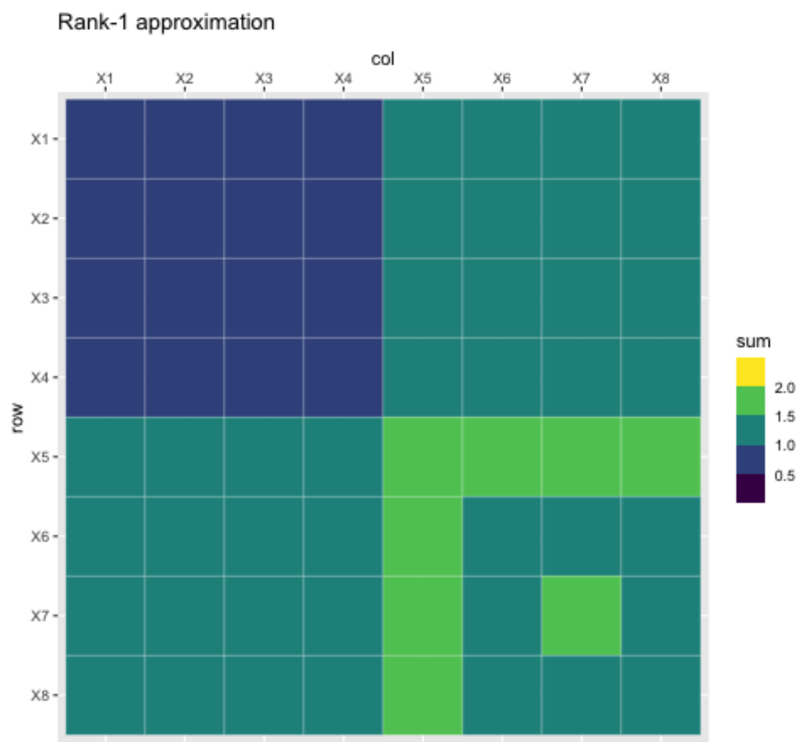
A little image



Rank-1 approximations



Increasing rank approximations



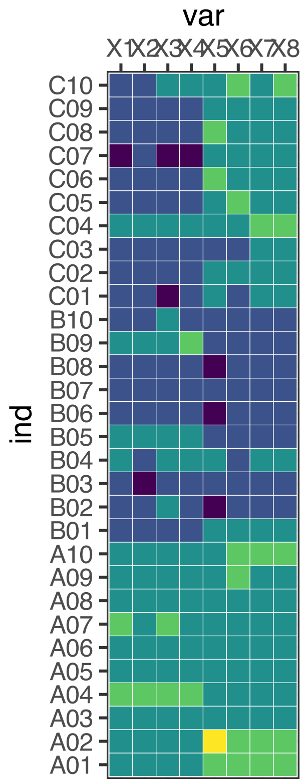
We can do the same kind of magic with the data itself

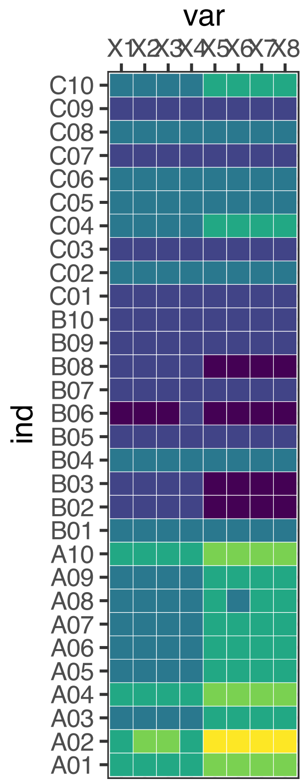
Singular value decomposition can be used to approximate a rectangular matrix with a lower ranked matrix of the same dimension

$$\arg\min_{\|\mathbf{u}\|_2^2=\|\mathbf{w}\|_2^2=1} \|\mathbf{X} - \delta \mathbf{u}\mathbf{w}^\top\|_F^2$$

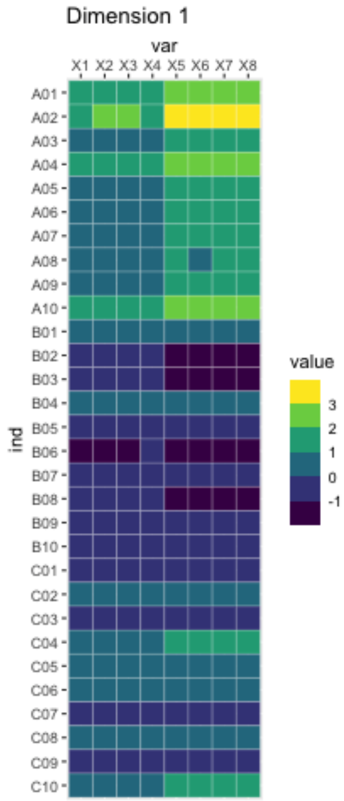
- δ: singular value
- $\mathbf{u}$: left singular vector
- $\mathbf{w}$: right singular vector

Rank 1 approximation

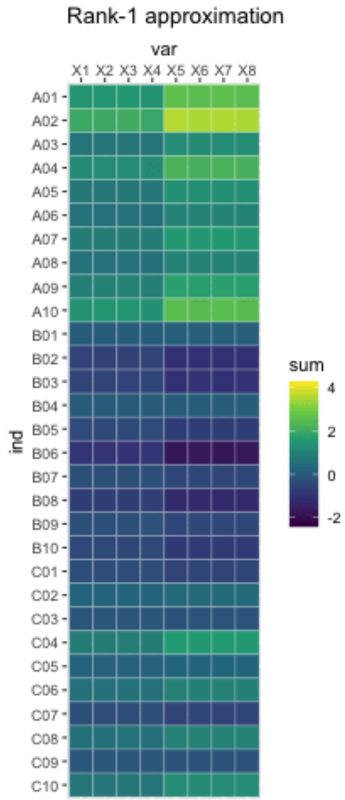




Rank-1 approximations



Increasing rank approximations



CONSTRAINING THE SVD

LASSO

LASSO is a (relatively) recent technique originally intended for regression problems:

$$\underset{\boldsymbol{\beta}}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 \text{ such that } \| \boldsymbol{\beta} \|_1 \leq r$$

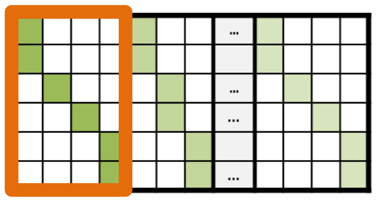
or the dual form

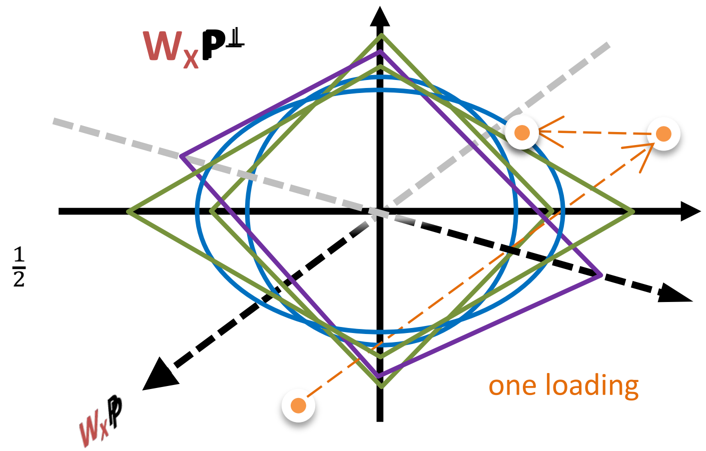
$$\underset{\boldsymbol{\beta}}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda \| \boldsymbol{\beta} \|_1$$

- the obtained weights are sparse (with zeros)
- the non-zeros coefficients correspond to important variables
- the result is biased
- selecting λ is done through cross-validation

Sparse GSVD (sGSVD) and sparse MCA (sMCA)

- Sparsify the GSVD: the sparse GSVD (sGSVD)
  - Intersection of 3 spaces
    - L_1-ball
    - L_2-ball with weights $\mathbf{P}^{\top}\mathbf{W}_{\mathbf{X}}\mathbf{P} = \mathbf{I}$
    - Orthogonal space:





- Sparse MCA (sMCA)
  - Generalize L_1-constraint
  - Group constraint:
    - $L_{(1,2)}$-norm: sum of L_2-norms
    - L_2-norms: [(Level 1 loading)^2 + (Level 2 loading)^2 + ⋯]

Optimization problem

$$(\delta_\ell, \mathbf{p}_\ell, \mathbf{q}_\ell) = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \|\mathbf{X} - \delta \mathbf{p}^\top \mathbf{q}\|_2^2$$

$$\text{such that } \begin{cases} \mathbf{p}^\top \mathbf{M} \mathbf{p} = \mathbf{q}^\top \mathbf{W} \mathbf{q} = 1 \\ \mathbf{p}^\top \mathbf{M} \mathbf{p}_{\ell'} = \mathbf{q}^\top \mathbf{W} \mathbf{q}_{\ell'} = 0, \forall \ell' < \ell \\ \| \mathbf{p} \|_1 \leq s_{\mathbf{p},\ell} \text{ and } \| \mathbf{q} \|_1 \leq s_{\mathbf{q},\ell} \end{cases}$$

Lost & Found

- Transition formulas: from rows to columns and back: pseudo-version

- Supplementary projection: pseudo-inverse projector

- Asymmetric projection (λ = 1): kept

- Distributional equivalence: kept

- Nested Solutions (i.e., $\mathbf{X}$ vs. $\mathbf{X} - \mathbf{rc}^{\mathrm{T}}$) and ($\lambda_1$ = 1): Lost

What are the parameters?

- $\mathbf{M}$ and $\mathbf{W}$ = masses, weights… so metrics
- $s_{\mathbf{p},\ell}$ and $s_{\mathbf{q},\ell}$ = sparsity parameters, between 1 (strong sparsity), and $\sqrt{\text{dimension}}$ (no sparsity)