Efficient dimension reduction

Given a collection of high dimensional vectors $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$, identify the efficient dimension m^* and the projector Π_{m^*} such that $\boldsymbol{Y}_i = \Pi_{m^*} \boldsymbol{Y}_i$ up to a measurement error for all i. The basic assumption is that the spectral gap $\lambda_{m^*} - \lambda_{m^*+1}$ is sufficiently large (larger that $\mathtt{C} n^{-1/2}$).

The procedure uses the idea of building a confidence set for the projector Π_m for each candidate efficient dimension m using the bootstrap procedure from Naumov, Spokoiny, Ulyanov (2017). The size of the confidence set is the crucial characteristics:

- for $m = m^*$, the eigen-subspace can be recovered with accuracy $1/\sqrt{n}$;
- for $m < m^*$, if $\lambda_m - \lambda_{m+1}$ is sufficiently big, again the confidence width is of order $n^{-1/2}$;
- for $m < m^*$ but $\lambda_m - \lambda_{m+1}$ is small or zero, then Π_m can be recovered up to an error within the e.d.r. subspace $\mathcal{I}$ of dimension m^*. This yields the confidence width of order m^*;
- if $m > m^*$, then the confidence width is of order $(m - m^*) * (d - m^*)$

Workpackages:

1. Efficient implementation working for large dimension d and big n
2. Exploring the theoretical properties: consistency and accuracy of estimation for the efficient dimension m^* and the corresponding eigen-subspace;
3. Application financial data
4. Application bio and medical data
5. Extension to a non-Gaussian case

Literature: Naumov, Spokoiny, Ulyanov (2017) Bootstrap confidence sets for spectral projectors of sample covariance. arXiv:1703.00871.

Contact: Alexey Naumov, Igor Silin, Alexander Podkopaev, VS