Robust Principal Component Analysis for Modal Decomposition of Corrupt Fluid Flows[2]

Project report by **Yerbol Palzhanov**, September 15

Motivation and challenges

Although experimental measuring techniques have evolved a lot over the past years, there are several well-known challenges to acquiring clean and accurate data. These include inadequate illumination and irregularities in the light field, background speckle, seeding density and nonpassivity of the particles, sharp gradients in flow properties, optical issues, such as alignment and aberration, limited resolution and shot noise in the image recording, and out of plane motion of the particles when measuring in 2D.

There are several different approaches to cleaning the data and extracting useful information form the data. This paper investigates the use of RPCA, a robust variant of POD/PCA.

RPCA was originally popularized with Netflix matrix competition and has since been widely used in different problems. In order to asses the performance of the RPCA, authors also do PCA and DMD modal analyses and compare results.

Brief outline of POD and DMD

- Proper orthogonal decomposition(POD)
- Dynamic mode decomposition (DMD)

TODO: put description of methods.

RPCA

Candes *et al.* [1] have developed RPCA that seeks to decompose a data matrix X into a structured low-rank matrix L that is characterized by dominant coherent structures and a sparse matrix S containing outliers and corrupt data:

$$X = L + S. \tag{1}$$

The principal components of L are robust to outliers and corrupt data, which are isolated in S.

Mathematically, the goal is to find $\mathbf{L}$ and $\mathbf{S}$ that satisfy the following:

$$\min_{\mathbf{L},\mathbf{S}} \operatorname{rank}(\mathbf{L}) + \|\mathbf{S}\|_0 \text{ subject to } \mathbf{L} + \mathbf{S} = \mathbf{X}.$$

$\|\mathbf{S}\|_0$ counts the number of nonzero elements in $\mathbf{S}$, quantifying how sparse it is. $\operatorname{rank}(\mathbf{L})$ is the number of nonzero singular values in $\mathbf{L}$, quantifying how many linearly independent rows and columns describe the data.

Since this is not a convex optimization problems, it is possible to solve the problem by convex relaxation. Relaxed convex problem is known as principal component pursuit and implemented in MATLAB using augmented Lagrange multiplier(ALM) algorithm.

Numerical experiments

RPCA filtering demonstrated on several data sets, drawn from direct numerical simulations and PIV data from experiments. We will focus on two of them: flow past a cylinder with artificial noise and cross-flow turbine experiment data from University of Washington.

Cylinder flow

Cross-flow turbine wake

TODO: Experiment with codes located at https://github.com/ischerl/RPCA-PIV

Bibliography

[1] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *J. ACM*, 58(3), jun 2011.

[2] I. Scherl, B. Strom, J. K. Shang, O. Williams, B. L. Polagye, and S. L. Brunton. Robust principal component analysis for modal decomposition of corrupt fluid flows. *Phys. Rev. Fluids*, 5:054401, May 2020.