ContactPose Dataset Analysis: Extraction of Grasping Synergies

Kyra Kerz

Abstract— **Grasping is natural for humans and their dexterity is far superior to that of robotic systems. Neuroscientific research reveals insights of human grasping behaviors, which can be transferred to robotic systems. The control of the human hand during grasping is dominated by movement in a low dimensional configuration space, compared to the higher degrees of freedom (DoF) the human hand possesses [1], [2]. To this purpose the ContactPose dataset, which is containing 2306 human hand postures of 50 subjects grasping 25 objects, is preprocessed and analyzed. Various dimensionality reduction methods are used to get a low level representation of the hand postures. The results are used to propose and evaluate different grasping synergies, which are ultimately transferred to the RBO Hand 3.**

I. Introduction

Controlling anthropomorphic robotic hands is a challenging task and controlling every joint independently is impractical due to many degrees of freedom and therefore an increasing complexity of the problem space. Transferring insight about human grasping behaviour can tackle this problem. The underlying principles include low dimensional representation of hand poses, meaning that far less control actions than number of joints are required to formulate a grasp. This is called hand synergies [2].

The goal of our project is to find a low dimensional representation of hand poses to reduce the dimension that is needed to control a robot hand. For this purpose the ContactPose dataset [3] is analyzed. It contains numerous recorded hand postures of people holding and using everyday objects. Corresponding joint coordinates are provided, which live in a higher dimensional space. As the coordinates are cartesian and therefore are dependent on orientation and placement in space, they need to be transformed in order to erase this dependency. For this purpose a kinematic model is created, which allows for an angle representation of the hand postures. Transforming from cartesian space to an angle representation is already a dimensionality reduction from 63 to 23 DoF. Further dimensionality reduction is achieved by using linear dimensionality reduction methods like principal component analysis (PCA) and independent component analysis (ICA).

The results for applying PCA on the dataset over all participant and object combinations are yielding two first principal components (PC) that capture almost two-third of the variance of all given hand postures. Those two PCs are used to recreate grasps as linear combination of those two components. Ultimately the results of this linear combinations are then deployed on the RBO Hand 3, showing that the calculated hand posture is able to grasp a variety of objects like a banana, lemon or wooden block.

For ICA its components can equally be used to recreate grasps as linear combination of various components. But as all of the components are equally important, there is no hierarchy on which the components are supposed to be chosen. Therefore three components that differ the most from each other, or in other words components that involve motions of different fingers, are chosen to recreate a grasp and the results are shown in simulation.

This work is structured as follows: first the related work is discussed in section II. After that the background of the used methods, as well as the dataset and its preprocessing are explained in section III. The next step is the description of the experiments that are done with the implemented methods and their results, which are explained in section IV. And finally a conclusion is given in section V.

II. Related Work

In this section the related work is discussed. One of the first studies for dimensionality reduction methods for grasping is conducted by Santello et al. [4]. They let five people reenact the grasping of 57 different objects and perform PCA on the measured grasps. Their results show that 80% of the variance could be explained with the first two principle components (PC). However in this case prescriptive grasps were analysed, but in reality fingers tend to be bent by the shape of the touched object and therefore some influences, for example compliance, aren't captured. Further more Ciocarlie et al. [5] are going further on [4] and are using the lower dimensional representation of the hand postures to formulate grasping synergies and applying them to various anthropomorphic robotic hands. Larger data sets for analyzing postural synergies are used by Nestor et al. [6]. In this study hand kinematic data of 77 subjects, performing up to 20 hand grasps of activities of daily living, was used. They found that 12 principal components were needed to achieve 80 % of variance. Also their work focuses more on grasping movements than single hand postures. Independent component analysis is a method for extracting the source signals of a mixed signal. It is often used on measurements like EEG signals which is presented by Sun et al. [7]. In this case ICA is used to separate different independent brain activities and therefore getting sources out of the mixed EEG data. Another use case for ICA in regards to images is proposed by Draper et al. [8]. They are using ICA for feature extraction on images. Instead of finding the most variance among faces, ICA aims to filter the independent components of each face. This results in extracting features like nose, eyes and eyebrows. A method for ICA being used on hand postures is presented by Kato et al. [9]. In their work they use

ICA basis vectors to represent local features that corresponds to actuation of singular fingers. But in their approach they are using time dependent video recording of finger movements compared to static hand postures that are provided by the ContactPose dataset. To my best of knowledge there are no studies so far that are using ICA on final hand posture observations.

III. Background

In this section the background and theoretical methods of our work are explained. First the kinematic model that is used to represent our data and its properties are elaborated. Second the theoretical background of the two methods used for dimensionality reduction are presented briefly.

A. ContactPose data set

For reaching the goal of recreating hand postures out of a lower dimensional representation the ContactPose data set is used. It provides hand-object contacts, paired with hand pose, object pose, and RGB-D images. ContactPose contains 2306 unique grasps that are gained by 50 participants who are able to grasp a selection of 25 3D-printed household objects in two different modi. The first modus is grasping an object as if the person would use it and the second modus is handing off the object. The contact between fingertips and the object is measured with a thermal camera. The thermal information about the position of the finger tips is combined with the hand postures of the participants that are recorded as multi-view RGB-D video clips. The images of the videos are then mapped to an object mesh using Open3D. The joint positions of the simulated 3D hand postures are given in cartesian coordinates.

B. Kinematic Model

The ContactPose dataset is providing the recorded hand postures in cartesian coordinates. This is not feasible as input for the dimensionality reduction methods, as cartesian coordinates capture different orientations and positions for each hand posture.

Therefore an angle representation for each joint is desired. To this purpose a kinematic model that can represent each hand posture is created. The kinematic model consists of 23 DoF. Those can be observed in figure 1, which shows all the joints responsible for the controllable flexion and abduction movements. The distal interphalangeal joint (DIP), proximal interphalangeal joint (PIP) and interphalangeal (IP) joints are represented by a one DoF hinge joint, that represents the flexion motion. This is also the case for the carpometacarpal (CMC) joints of the ring finger and pinkie as those are need in order to recreate the arch of the back of the hand. For each of the metacarpophalangeal (MCP) joints a 2 DoF universal joint, consisting of two hinge joints, is used in order to mimic the motion of abduction and flexion. While the 23 angles representing the 23 DoF as input for the dimensionality reduction methods, 6 more angles are needed in order to fully construct the hand. Two angles for abduction and flexion are measure at the Thumb base, and four for the

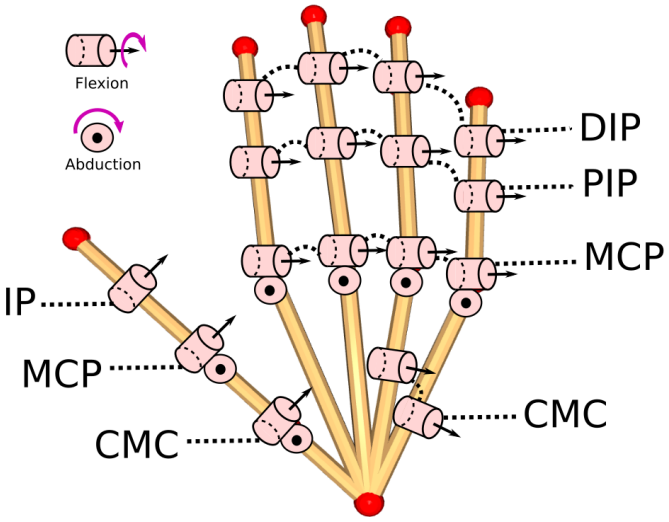


Fig. 1: Kinematic model with 23 DoF

abduction of the CMC joints of the other four fingers. Those angles are calculated as mean over all the hand shapes of all participants.

C. Principal Component Analysis

Principal Component Analysis (PCA) is a linear dimensionality reduction method for feature extraction. It gives the directions that maximize the variance of the data, which are called principal components (PC). Out of the mean centered observations for the hand postures the covariance matrix is computed. The covariance matrix can then be used to calculate its eigenvectors and eigenvalues. The eigenvectors or principal components represent the direction of determine the directions of the new feature space, and the eigenvalues determine the magnitude.
Hand postures can now be recreated by summing up the principal components which are scaled with a coefficient and then being added to the mean, which is described by equation 1. In most works [4], [5] the two first principal components are used to recreate the grasps, as they are capturing the most variance.

$$\text{Grasp} = mean + \sum_{i=1}^{\#PC} c_i \cdot PC_i \tag{1}$$

The range of the coefficients is determined by the covariance between the original data and the unit-scaled components. For the implementation the PCA class of the scikit-learn library is used [10].

D. Independent Component Analysis

Independent Component Analysis (ICA) is a method for separating linear mixed signals into source signals or components. The goal for the source signals is to be statistically independent or in other words minimization of mutual information [11]. Suppose there is random vector $\mathbf{x}$ which consists of $n \in \mathbb{N}$ observations $x_1(m), ..., x_n(m)$ with $m \in \mathbb{N}$ features. In the setting for ICA these observations are modeled as

	Finger 1-4
MCP Abduction	$-15° \leq \theta \leq 15°$
MCP Flexion	$0° \leq \theta \leq 90°$
PIP Flexion	$0° \leq \theta \leq 90°$
DIP Flexion	$0° \leq \theta \leq 110°$

TABLE I: Constraints for each joint and its direction of motion for all fingers except the thumb

linear combination of n source signals $s_1(m),...,s_g(m)$ with $g \in \mathbb{N}$, which are forming a latent source vector $\mathbf{s}$. Also g as the number of source signals The generative model can be expressed as:

$$\mathbf{x} = \sum_{i=1}^{n} \mathbf{a}_i s_i \tag{2}$$

In this case $\mathbf{a}_i$ is denoting the columns of the mixing matrix $\mathbf{A}$. The independent components of a mixed signal can be recovered by the inverse of $\mathbf{A}$ which is the unmixing Matrix $\mathbf{W}$:

$$\mathbf{s} = \mathbf{W}\mathbf{x} \tag{3}$$

As the source signals and mixing matrix are unknown, any scalar multiplier of one of the source signals can also divide the corresponding column of the mixing matrix by that scalar. This means that the sign as well as the magnitude are ambiguous. Furthermore there is no order of importance for the independent components, all components are equally important [11]. For ICA the grasps are similar recreated like for the PCA. A chosen amount of source signals, which are each scaled with a coefficient, then summed up and then added to the mean grasp:

$$\text{Grasp} \ = mean + \sum_{i=1}^{\#S} c_i \cdot S_i \tag{4}$$

As the sign and amplitude of the source signal are ambiguous, the coefficients need to scale the values of the angles so that they lie within the constraints of the joint movements of a human hand that are presented in [12]. An overview of the joint limits is given in table 1. The table shows the constraints for each joint of four fingers except for the thumb. For the thumb flexion and abduction for CMC and flexion of the IP lie between $0° \leq \theta \leq 90°$ each, while the constraints for the MCP abduction and flexion are chosen similar to the other fingers listed in table 1.

For the implementation the FastICA class of the scikit-learn library is used [13].

IV. Experiments & Results

In this section the results of the implementations described in III are discussed. First the experiment results of the PCA analysis are presented and after that the simulation results of the ICA are shown. Finally the results of the PCA analysis are deployed on the RBO Hand 3.

A. Results of PCA

In order to get a low dimensional representation of the data PCA is applied. The goal is to be able to recreate a grasp with just a few principal components instead of actuating each joint individually. This means the number of components has to be a lot lower than the number of DoF of the hand postures. As a lot of recent studies are focusing on the first two principal components, like Santello et al. [4], this experiment will also try to replicate a grasp out of the first two principal components. For this purpose the observations of hand postures of the ContactPose data set, which contains 50 and 25 objects in the two modi *use* and *handoff*, are used. As there aren't all combinations of object, person and modi as usable data for the hand postures provided, overall 1947 hand postures as input for the analysis are left. The results can be seen in figure 2.

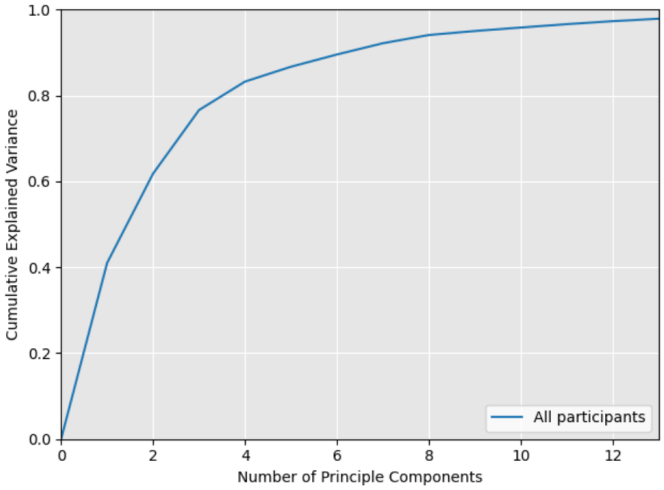


Fig. 2: Cumulative Variance over 12 principal components for 1947 hand posture observations

It displays the cumulative variance on the y-axis for each component depicted on the x-axis. Overall 12 principal components are needed to achieve 100% variance. The first two components account for 62% of the variance. With the first component accounting for 41% of the variance. Out of the first two principle components it is possible to recreate a grasp. This is done with the help of equation (1) that allows for a linear combination of the principle components. A visualisation of those linear combinations can be observed in figure 3. In the middle is the mean grasp over all hand postures to be seen. On the horizontal axis the linear combination of the first PC together with the mean grasp can be observed. For an increasing coefficient the fingers are bending in direction of the palm, while a decreasing of the coefficient leads to the fingers rolling out to a relative flat hand. The range of the coefficients for the first and second principal component lies between -1 and 1. For the second PC and a coefficient close to 1, the abduction angle of the CMC joint of the thumb is getting bigger, leading to the thumb moving closer to the index finger and the grasp is

closing. For a smaller coefficient the result can be seen in the bottom of the figure. The grasp is opening, especially the index finger and thumb.

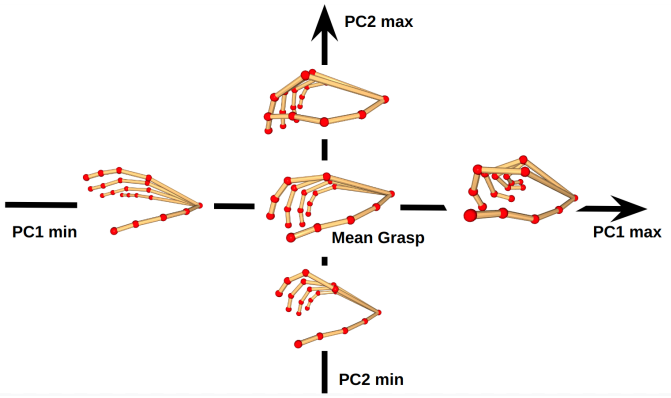


Fig. 3: Overview of the linear combinations for PC1 and PC2

The result of the linear combination of both PCs can be observed in figure 4. Again the mean grasp is observable in the middle of the plot. It already shows a bigger actuation of the pinky and ring finger compared to the middle and index finger. For a smaller coefficient the hand is opening to a flat hand shape. For a bigger coefficient the grasp is closing. As it is already observable for each PC individually the thumb is mainly actuating in its abduction motion. Furthermore the index finger isn't as much actuated as the middle finger, ring finger and pinky, which is also already observable in figure 4.

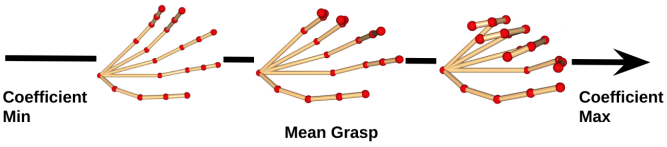
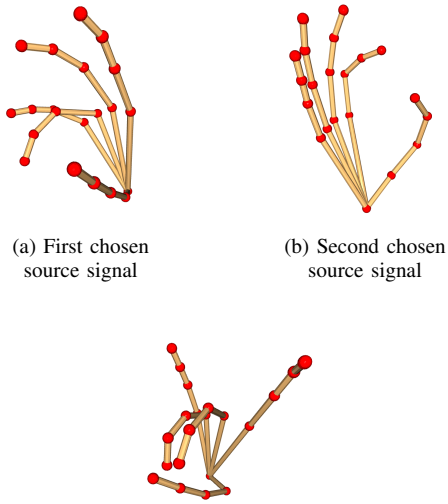


Fig. 4: Linear combination of the combined PC1 and PC2

B. *Results of ICA*

As described in the related work section II, there is no similar work done so far where ICA is applied on a dataset with descriptive hand poses. As the number of source signals is at maximum equal to the number of observations, a subset of the data is chosen. 40 observations of each participant grasping an apple are analyzed. This is done, because it makes it easier to understand the results yielded by the ICA over hand postures, as there is some intuition about how the actuation of the fingers are supposed to look like. For the ICA also three components were chosen, as three components were the minimum of components that showed actuation of all fingers. As there is no indication about which components are supposed to be chosen, because all of them are equally important [11], we are looking through all the 40 components in order to find three that are able to formulate a grasp. For this purpose in all of the three chosen components of



(a) First chosen source signal

(b) Second chosen source signal

(c) Third chosen source signal

Fig. 5: Visualized results of the ICA. The images a) - c) are showing each of the source signals individually

ICA, different fingers are actuated. Equation (4) is used to formulate the grasp out of the source signals.

Figure 5 is showing the results of our findings. Image a) shows a component that has a high actuation of the pinky and ring finger. Image c) shows a hand posture with focus of the bending of the index finger for encapsulating a round object. Choosing those three components ensure that all fingers are actuated and part of the hand posture for the grasping. The linear combination of the three chosen source hand postures as well as the mean can be observed in figure 6. It shows the mean grasp in the middle. Left of the mean grasp is the linear combination of the hand postures with a very small coefficient depicted. The chosen coefficients range from -0.2 up to 0.2, as those values don't violate the constraints which are shown in table 1.

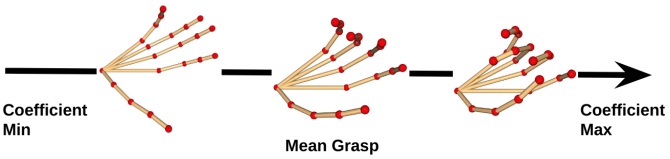
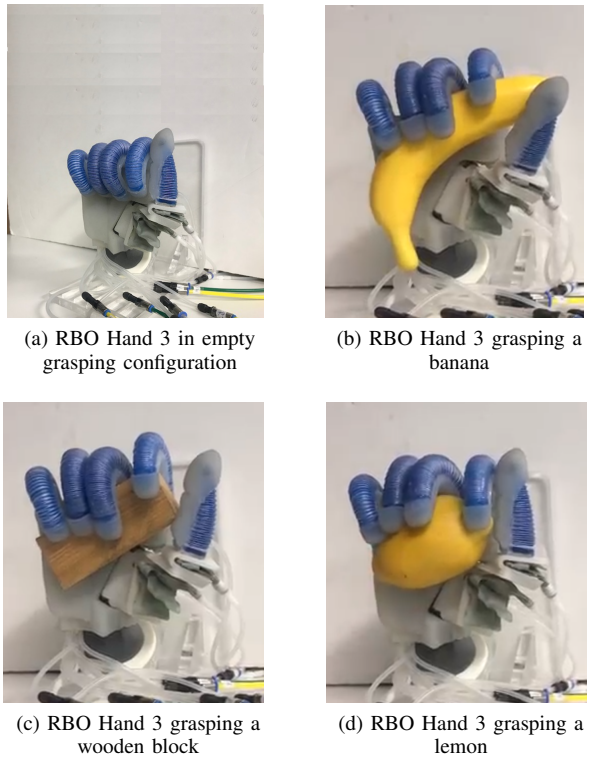


Fig. 6: Linear combination of the three source signals

For a larger coefficient the hand posture on the right is closing. Due to two single source components having influence on the ring finger, it gets actuated more intensively. But compared to the results of the PCA in this case the thumb is also a lot more actuated and involved in the grasping movement. This again shows that is important to select the components complement each other the best, even though in theory all components are equally important.

C. Results on RBO Hand 3

The first two principle components and the mean grasp of the experiments of the PCA described in section IV are deployed on the RBO Hand 3. To this purpose the configuration with all angles set to zero of the kinematic model has the longitudinal axes of the index finger and the longitudinal axes of the thumb being perpendicular to each other. In order to map the angles of the recreated grasp to the RBO Hand 3, the angles have to be normalized to 90 degrees. The results can be observed in figure 7. Image a) shows the closed grasp of the applied PCs. All four fingers are closed evenly, with the thumb being adjacent to the index finger. Overall the thumb isn't actuated much as the variance of first two PCs isn't covering enough information about the thumb. Choosing more components could counter this problem. For lengthy objects like bananas it isn't a problem as can be observed in image b), as there is an abduction motion of the thumb. Also the data set inhibit a range of lengthy objects like banana, bottle or flashlight [3]. Image c) shows the grasping of a smaller object, a wooden block. In this case a larger angle for flexion of the CMC joint would result in a more natural kind of grasp, even thought the other fingers are able to grasp the block. The same observation can be made for the scenario shown in image d). The RBO Hand 3 is able to grasp the lemon, but this is not the natural grasping as humans would do it, as the thumb is not much actuated.



(a) RBO Hand 3 in empty grasping configuration

(b) RBO Hand 3 grasping a banana

(c) RBO Hand 3 grasping a wooden block

(d) RBO Hand 3 grasping a lemon

Fig. 7: Results of the recreated grasp using PC1 and PC2 deployed on the RBO Hand 3. The images a) - d) are showing the empty grasp and grasping of the objects banana, wooden block and lemon respectively

V. Conclusion

In this work 2306 hand postures given by the ContactPose dataset were analyzed. After transforming the hand postures from a cartesian space to an angle representation, dimensionality reduction methods like PCA and ICA were used in order to get a lower representation of the grasping space. Finally the results of the PCA could be applied an the RBO Hand 3. We showed that insights of human grasping behaviour could indeed be transferred to a robotic system: the control of the human hand during grasping is dominated by movement in a low dimensional configuration space, compared to the higher degrees of freedom (DoF) that the human hand possesses [1], [2]. For further improvements of the results could be achieved by using non-linear dimensionality reduction tools like kernel-PCA or autoencoders. Also for ICA a more sophisticated method like kurtosis for ranking the components in order to formulate a grasp could be used as described in [14].

References

[1] M. Schieber, "Muscular production of individuated finger movements: the roles of extrinsic finger muscles," *Journal of Neuroscience*, vol. 15, no. 1, pp. 284–297, 1995. [Online]. Available: https://www.jneurosci.org/content/15/1/284

[2] G. Torres-Oviedo and L. Ting, "Muscle synergies characterizing human postural responses contact information," 2007.

[3] S. Brahmbhatt, C. Tang, C. D. Twigg, C. C. Kemp, and J. Hays, "Contactpose: A dataset of grasps with object contact and hand pose," 2020.

[4] M. Santello, M. Flanders, and J. Soechting, "Postural hand synergies for tool use," *The Journal of Neuroscience*, vol. 18, pp. 10 105 – 10 115, 1998.

[5] M. Ciocarlie, C. Goldfeder, and P. Allen, "Dexterous grasping via eigengrasps : A low-dimensional approach to a high-complexity problem," 2007.

[6] N. Jarque-Bou, A. Scano, M. Atzori, and H. Müller, "Kinematic synergies of hand grasps: a comprehensive study on a large publicly available dataset," *Journal of NeuroEngineering and Rehabilitation*, vol. 16, 05 2019.

[7] Lisha Sun, Ying Liu, and P. J. Beadle, "Independent component analysis of eeg signals," in *Proceedings of 2005 IEEE International Workshop on VLSI Design and Video Technology, 2005.*, 2005, pp. 219–222.

[8] B. A. Draper, K. Baek, M. S. Bartlett, and J. Beveridge, "Recognizing faces with pca and ica," *Computer Vision and Image Understanding*, vol. 91, no. 1, pp. 115–137, 2003, special Issue on Face Recognition. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1077314203000778

[9] K. Makoto, C. Yen-Wei, and X. Gang, "Articulated hand motion tracking using ica-based motion analysis and particle filtering," *Journal of Multimedia*, vol. 1, 06 2006.

[10] "Sklearn PCA," https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html, accessed: 2021-02-22.

[11] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Networks*, vol. 13, no. 4, pp. 411–430, 2000. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0893608000000265

[12] John Lin, Ying Wu, and T. S. Huang, "Modeling the constraints of human hand motion," in *Proceedings Workshop on Human Motion*, 2000, pp. 121–126.

[13] "Sklearn FastICA," https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.FastICA.html#sklearn-decomposition-fastica, accessed: 2021-02-22.

[14] G. Brys, M. Hubert, and A. Struyf, "Robust measures of tail weight," *Computational Statistics & Data Analysis*, vol. 50, no. 3, pp. 733–759, 2006. [Online]. Available: https://www.sciencedirect.