ECKO: Ensemble of Clustered Knockoffs for robust multivariate inference on MRI data

Tuan-Binh Nguyen, Jérôme-Alexis Chevalier, Bertrand Thirion

ECKO: Ensemble of Clustered Knockoffs for robust multivariate inference on MRI data

Tuan-Binh Nguyen[1,2,3], Jérôme-Alexis Chevalier[1,2,4], and Bertrand Thirion[1,2]

[1] Parietal team, Inria Saclay, France
[2] CEA/Neurospin, Gif-Sur-Yvette, France
[3] LMO - Laboratoire de Mathématiques d'Orsay
[4] Telecom Paristech, Paris, France

Abstract. Continuous improvement in medical imaging techniques allows the acquisition of higher-resolution images. When these are used in a predictive setting, a greater number of explanatory variables are potentially related to the dependent variable (the response). Meanwhile, the number of acquisitions per experiment remains limited. In such high dimension/small sample size setting, it is desirable to find the explanatory variables that are truly related to the response while controlling the rate of false discoveries. To achieve this goal, novel multivariate inference procedures, such as knockoff inference, have been proposed recently. However, they require the feature covariance to be well-defined, which is impossible in high-dimensional settings. In this paper, we propose a new algorithm, called Ensemble of Clustered Knockoffs, that allows to select explanatory variables while controlling the false discovery rate (FDR), up to a prescribed spatial tolerance. The core idea is that knockoff-based inference can be applied on groups (clusters) of voxels, which drastically reduces the problem's dimension; an ensembling step then removes the dependence on a fixed clustering and stabilizes the results. We benchmark this algorithm and other FDR-controlling methods on brain imaging datasets and observe empirical gains in sensitivity, while the false discovery rate is controlled at the nominal level.

1 Introduction

Medical images are increasingly used in predictive settings, in which one wants to classify patients into disease categories or predict some outcomes of interest. Besides predictive accuracy, a fundamental question is that of *opening the black box*, *i.e.* understanding the combinations of observations that explains the outcome. A particular relevant question is that of the importance of image features in the prediction of an outcome of interest, conditioned on other features. Such conditional analysis is a fundamental step to allow causal inference on the implications of the signals from image regions in this outcome; see e.g. [12] for the case of brain imaging. However, the typical setting in medical imaging is that of high-dimensional small-sample problems, in which the number of samples n is much smaller than the number of covariates p. This is further aggravated by the

steady improvements in data resolution. In such cases, classical inference tools fail, both theoretically and practically. One solution to this problem is to reduce the massive number of covariates by utilizing dimension reduction, such as clustering-based image compression, to reduce the number of features to a value close to n; see e.g. [4]. This approach can be viewed as the bias/variance trade-off: some loss in the localization of the predictive features —bias— is tolerated as it comes with less variance —hence higher power— in the statistical model. This is particularly relevant in medical imaging, where localizing predictive features at the voxel level is rarely important: one is typically more interested in the enclosing region.

However, such a method suffers from the arbitrariness of the clustering step and the ensuing high-variance in inference results with different clustering runs, as shown empirically in [6]. [6] also introduced an algorithm called Ensemble of Clustered Desparsified Lasso (ECDL), based on the inference technique developed in [13], that provides p-values for each feature, and controls the Family Wise Error Rate (FWER), i.e. the probability of making one or more false discoveries. In applications, it is however more relevant to control the False Discovery Rate (FDR) [3], which indicates the expected fraction of false discoveries among all discoveries, since it allows to detect a greater number of variables. In univariate settings, the FDR is easily controlled by the Benjamini-Hochberg procedure [3], valid under independence or positive correlation between features. It is unclear whether this can be applied to multivariate statistical settings. A promising method which controls the FDR in multivariate settings is the so-called knockoff inference [2, 5], which has been successfully applied in settings where $n \approx p$. However, the method relies on randomly constructed knockoff variables, therefore it also suffers from instability. Our contribution is a new algorithm, called Ensemble of Clustered Knockoffs (ECKO), that *i)* stabilizes knockoff inference through an aggregation approach; *ii)* adapts knockoffs to $n \ll p$ settings. This is achieved by running the knockoff inference on the reduced data and ensembling the ensuing results.

The remainder of our paper is organized as follows: Sec. 2 establishes a rigorous theoretical framework for the ECKO algorithm; Sec. 3 describes the setup of our experiments with both synthetic and brain imaging data predictive problems, to illustrate the performance of ECKO, followed by details of the experimental results in Sec. 4; specifically, we benchmark this approach against the procedure proposed in [7], that does not require the clustering step, yet only provides asymptotic ($n \to \infty$) guarantees. We show the benefit of the ECKO approach in terms of both statistical control and statistical power.

2 Theory

2.1 Generalized Linear Models and High Dimensional Setting

Given a design matrix $\mathbf{X} \in \mathbb{R}^{n \times p}$ and a response vector $\mathbf{y} \in \mathbb{R}^n$, we consider that the true underlying model is of the following form:

$$\mathbf{y} = f(\mathbf{X}\mathbf{w}^*) + \sigma\boldsymbol{\epsilon} \ , \tag{1}$$

where $\mathbf{w}^* \in \mathbb{R}^p$ is the true parameter vector, $\sigma \in \mathbb{R}^+$ the (unknown) noise magnitude, $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$ the noise vector and f is a function that depends on the experimental setting (e.g. $f = Id$ for the regression problem or e.g. $f = sign$ for the classification problem). The columns of $\mathbf{X}$ refer to the explanatory variables also called features, while the rows of $\mathbf{X}$ represent the coordinates of different samples in the feature space. We focus on experimental settings in which the number of features p is much greater than the number of samples n *i.e.* $p \gg n$. Additionally, the (true) support denoted by S is given by $S = \{k \in [p] : \mathbf{w}^*_k \neq 0\}$. Let $\hat{S}$ denotes an estimate of the support given a particular inference procedure. We also define the signal-to-noise ratio (SNR) which allows to assess the noise regime of a given experiment:

$$\text{SNR} = \frac{\|\mathbf{X}\mathbf{w}^*\|_2^2}{\sigma^2 \|\boldsymbol{\epsilon}\|_2^2} \ . \tag{2}$$

A high SNR means the signal magnitude is strong compared to the noise, hence it refers to an easier inference problem.

2.2 Structured Data

In medical imaging and many other experimental settings, the data stored in the design matrix $\mathbf{X}$ relate to *structured* signals. More precisely, the features have a peculiar dependence structure that is related to an underlying spatial organization, for instance the spatial neighborhood in 3D images. Then, the features are generated from a random process acting on this underlying metric space. In our paper, the distance between the j-th and the k-th features is denoted by $d(j, k)$.

2.3 FDR control

In this section, we introduce the false discovery rate (FDR) and a spatial generalization of the FDR that we called δ-FDR. This quantity is important since a desirable property of an inference procedure is to control the FDR or the δ-FDR. In the following, we assume that the true model is the one defined in Sec. 2.1.

Definition 1 False discovery proportion (FDP). *Given an estimate of the support $\hat{S}$ obtained from a particular inference procedure, the false discovery proportion is the ratio of the number selected features that do not belong to the support (false discoveries) divided by the number of selected features (discoveries):*

$$FDP = \frac{\#\{k \in \hat{S} : k \notin S\}}{\#\{k \in \hat{S}\}} \tag{3}$$

Definition 2 δ-FDP. *Given an estimate of the support $\hat{S}$ obtained from a particular inference procedure, the false discovery proportion with parameter $\delta > 0$, denoted δ-FDP is the ratio of the number selected features that are at a distance*

more than δ from any feature of the support, divided by the number of selected features:

$$\delta\text{-}FDP = \frac{\#\{k \in \hat{S} : \forall j \in S,\ d(j,k) > \delta\}}{\#\{k \in \hat{S}\}} \tag{4}$$

One can notice that for $\delta = 0$, the FDP and the δ-FDP refer to same quantity *i.e.* $0\text{-}FDP = FDP$.

Definition 3 False Discovery Rate (FDR) and δ-FDR. *The false discovery rate and the false discovery rate with parameter $\delta > 0$ which is denoted by δ-FDR are respectively the expectations of the FDP and the δ-FDP:*

$$\begin{aligned} FDR &= \mathbb{E}[FDP]\ , \\ \delta\text{-}FDR &= \mathbb{E}[\delta\text{-}FDP]\ . \end{aligned} \tag{5}$$

2.4 Knockoff Inference

Initially introduced by [2] to identify variables in genomics, the knockoff filter is an FDP control approach for multivariate models. This method has been improved to work with mildly high-dimensional settings in [5], leading to the so-called model-X knockoffs:

Definition 4 Model-X knockoffs [5]. *The model-X knockoffs for the family of random variables $\mathbf{X} = (X_1, \dots X_p)$ are a new family of random variables $\tilde{\mathbf{X}} = (\tilde{X}_1, \dots, \tilde{X}_p)$ constructed to satisfy the two properties:*

1. *For any subset $\mathcal{K} \subset \{1, \dots, p\}$: $(\mathbf{X}, \tilde{\mathbf{X}})_{swap(\mathcal{K})} \overset{d}{=} (\mathbf{X}, \tilde{\mathbf{X}})$,*
   where the vector $(\mathbf{X}, \tilde{\mathbf{X}})_{swap(\mathcal{K})}$ denotes the swap of entries X_j and $\tilde{X}_j$, $\forall j \in \mathcal{K}$
2. *$\tilde{\mathbf{X}} \perp\!\!\!\perp \mathbf{y} \mid \mathbf{X}$ where $\mathbf{y}$ is the response vector.*

In a nutshell, knockoff procedure first creates extra null variables that have a correlation structure similar to that of the original variables. A test statistic vector is then calculated to measure the strength of the original versus its knockoff counterpart. An example of such statistic is the lasso-coefficient difference (LCD) that we use in this paper:

Definition 5 Knockoff procedure with intermediate p-values [2, 5].

1. *Construct knockoff variables, produce matrix concatenation: $[\mathbf{X}, \tilde{\mathbf{X}}] \in \mathbb{R}^{n \times 2p}$.*
2. *Calculate LCD by solving*

$$\min_{\mathbf{w} \in \mathbb{R}^{2p}} \frac{1}{2} \|\mathbf{y} - [\mathbf{X}, \tilde{\mathbf{X}}]\mathbf{w}\|_2^2 + \lambda \|\mathbf{w}\|_1\ ,$$

   and then, for all $j \in [p]$, take the difference $z_j = |\hat{\mathbf{w}}_j(\lambda)| - |\hat{\mathbf{w}}_{j+p}(\lambda)|$.
3. *Compute the p-values p_j, for $j \in [p]$:*

$$p_j = \frac{\#\{k : z_k \leq -z_j\}}{p}\ . \tag{6}$$

4. *Derive q-values by Benjamini-Hochberg procedure:* $(q_j)_{j\in[p]} = BHq\left((p_j)_{j\in[p]}\right)$
5. *Given a desired FDR level* $\alpha \in (0,1)$*:* $\hat{S} = \{j : q_j \leq \alpha\}$.

Remark 1 *The above formulation is distinct from that of [2, 5], but it is equivalent. We use it to introduce the intermediate variables* p_j *for all* $j \in [p]$.

Our first contribution is to extend this procedure computing q_j by aggregating different draws of knockoffs before applying the Benjamini-Hochberg (BHq) procedure. More precisely, we first compute B draws of knockoff variables and, using (6), we derive the corresponding p-values $p_j^{(b)}$, for all $j \in [p]$ and $b \in [B]$. Then, we aggregate them for each j in parallel, using the quantile aggregation procedure introduced in [11]:

$$\forall j \in [p],\ p_j = \text{quantile-aggregation}(\{p_j^{(b)} : b \in [B]\}) \tag{7}$$

We then proceed with the fourth and fifth steps of the knockoff procedure described in Def. 5.

2.5 Dimension reduction

Knockoff (KO) inference is intractable in high-dimensional settings, as knockoff generation requires the estimation and inversion of covariance matrices of size $(2p \times 2p)$. Hence we leverage data structure by introducing a clustering step that reduces data dimension before applying KO inference. As in [6], assuming the features' signals are spatially smooth, it is relevant to consider a spatially-constrained clustering algorithm. By averaging the features with each clustering, we reduce the number of parameters from p to q, the number of clusters, where $q \ll p$. KO inference on cluster-based signal averages will be referred to as Clustered Knockoffs (CKO). However, it is preferable not to fully rely on a particular clustering, as a small perturbation on the input data has a dramatic impact on the clustering solution. We followed the approach used in [9] that aggregates solutions across random clusterings. More precisely, they build C different clusterings from C different random subsamples of size $\lfloor 0.7n \rfloor$ from the full sample $\mathbf{X}$, but always using the same clustering algorithm.

2.6 The Ensemble of Clustered Knockoff Algorithm

The problem is to aggregate the q-values obtained across CKO runs on different clustering solutions. To do so, we transfer the q-values from clusters (group of voxels) to features (voxels): given a clustering solution $c \in [C]$, we assign to each voxel the q-value of its corresponding cluster. More formally, if, considering the c-th clustering solution, the k-th voxel belongs to the j-th cluster denoted by $G_j^{(c)}$ then the q-value $\tilde{q}_k^{(c)}$ assigned to this voxel is: $\tilde{q}_k^{(c)} = q_j^{(c)}$ if $k \in G_j^{(c)}$. This procedure hinges on the observation that the FDR is a resolution-invariant concept —it controls the *proportion* of false discoveries. In the worst case, this results in a spatial inaccuracy of δ in the location of significant activity, δ being

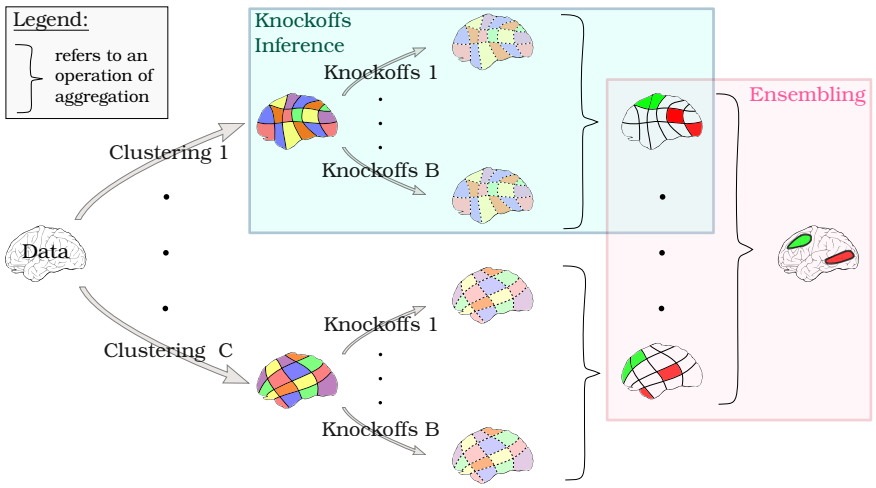


Fig. 1: Representation of the ECKO algorithm. To create a stable inference result, we introduce ensembling steps both within each cluster level and at the voxel-level, across clusterings.

the diameter of the clusters. Finally, the aggregated q-value $\tilde{q}_k$ of the k-th voxel is the average of the q-values $\tilde{q}_k^{(c)}$, $c \in [C]$, received across C different clusterings: given the FDR definition (5), FDPs can naturally be averaged. The algorithm is summarized in Alg. 1 and represented graphically in Fig. 1.

2.7 Theoretical Results

Ensemble of Clustered Knockoffs (ECKO).

Theorem 1 δ-FDR control by the Ensemble of Clustered Knockoffs procedure at the voxel level. *Assuming that the true model is the one defined in (1), using the q-values $\tilde{q}_k$ defined in Sec. 2.6, the estimated support $\hat{S} = \{k : \tilde{q}_k \leq \alpha\}$ ensure that the δ-FDR is lower than α.*

Sketch of the proof (details are omitted for the sake of space). We first establish that the aggregation procedure yields q-values q_j^c that control the FDR. This follows simply from the argument given in the proof of Theorem 3.1 in [11]. Second, we show that broadcasting the values from clusters (q) to voxels ($\tilde{q}$) still controls the FDR, yet with a possible inaccuracy of δ, where δ is the supremum of clusters diameters: the δ-FDR is controlled. This comes from the resolution invariance of FDR and the definition of δ-FDR. Third, averaging-based aggregation of the q-values at the voxel level, controls the δ-FDR. This stems from the definition of the FDR as an expected value.

Algorithm 1: Full ECKO algorithm

input : Data matrix $\mathbf{X}_{\texttt{init}} \in \mathbb{R}^{n \times p}$, response vector $\mathbf{y}_{\texttt{init}} \in \mathbb{R}^n$; Clustering object $\texttt{Ward}(\cdot)$;

param : $q = 500, B = 25, C = 25$, `fdr` - Nominal FDR threshold;

for $c = 1, \ldots, C$ **do**

$X^{(c)}_{\texttt{init}} = \texttt{resample}(X_{\texttt{init}})$

$X^{(c)}_{\texttt{clustered}} = \texttt{Ward}(q, X^{(c)}_{\texttt{init}})$

for $b = 1, \ldots, B$ **do**

$\forall j = 1, \ldots, q :$

$z_j^{(b,c)} \leftarrow \texttt{Knockoffs}(X^{(c)}_{\texttt{clustered}}, \mathbf{y}_{\texttt{init}}, \texttt{fdr})$

$p_j^{(b,c)} \leftarrow \dfrac{\#\{k \in [q] : z_k^{(b,c)} \leq -z_j^{(b,c)}\}}{p}$

end

$\forall j = 1, \ldots, q :$

$p_j^{(c)} \leftarrow \texttt{Aggregated}(p_j^{(b,c)}, b \in [B])$

$q_j^{(c)} \leftarrow \texttt{BHq\_corrected}(p_j^{(c)})$

$\forall k = 1, \ldots, p :$

$\tilde{q}_k^{(c)} \leftarrow q_j^{(c)}$ if $k \in G_j^{(c)}$

end

$\forall k = 1, \ldots, p :$

$\tilde{q}_k \leftarrow \texttt{Average}(\tilde{q}_k^{(c)}, c \in [C])$

return $\hat{S} \leftarrow \{k \in [p] : \tilde{q}_k \leq \texttt{fdr}\}$

2.8 Alternative approaches

In the present work, we use two alternatives to the proposed CKO/ECKO approach: the ensemble of clustered desparsified lasso (ECDL) [6] and the APT framework from [7]. As we already noted, ECDL is structured as ECKO. The main differences are that it relies on desparsified lasso rather than knockoff inference and returns p-values instead of q-values. The APT approach was proposed to return feature-level p-values for binary classification problems (though the generalization to regression is straightforward). It directly works at the voxel level, yet with two caveats:

- Statistical control is granted only in the $n \to \infty$ limit
- Unlike ECDL and ECKO, it is unclear whether the returned score represents marginal or conditional association of the input features with the output.

For both ECDL and APT, the returned p-values are converted to q-values using the standard BHq procedure. The resulting q-values are questionable, given that BHq is not valid under negative dependence between the input q-values [3]; on the other hand, practitioners rarely check the hypothesis underlying statistical models. We thus use the procedure in a black-box mode and check its validity a posteriori.

3 Experiments

Synthetic data. To demonstrate the improvement of the proposed algorithm, we first benchmark the method on 3D synthetic data set that resembles a medical image with compact regions of interest that display some predictive information. The size of the weight vector $\mathbf{w}$ is $50 \times 50 \times 50$, with 5 regions of interest (ROIs) of size $6 \times 6 \times 6$. A design matrix $\mathbf{X}$ that represents random brain signal is then sampled according to a multivariate Gaussian distribution. Finally, the response vector $\mathbf{y}$ is calculated following linear model assumption with Gaussian noise, which is configured to have $SNR \approx 3.6$, similar to real data settings. An average precision-recall curve of 30 simulations is calculated to show the relative merits of single cluster Knockoffs inference versus ECKO and ECDL and APT. Furthermore, we also vary the Signal-to-Noise Ratio (SNR) of the simulation to investigate the accuracy of FDR control of ECKO with different levels of difficulty in detecting the signal.

Real MRI dataset. We compare single-clustered Knockoffs (CKO), ECKO and ECDL on different MRI datasets downloaded from the Nilearn library [1]. In particular, the following datasets are used:

- **Haxby** [8]. In this functional-MRI (fMRI) dataset, subjects are presented with images of different objects. For the benchmark in our study, we only use the brain signal and responses for images related to faces and houses of subject 2 ($n = 216, p = 24083$).
- **Oasis** [10]. The original collection include data of gray and white matter density probability maps for 416 subjects aged 18 to 96, 100 of which have been clinically diagnosed with very mild to moderate Alzheimers disease. The purpose for our inference task is to find regions that predict the age of a subject ($n = 400, p = 153809$).

We chose $q = 500$ in all experiments for the algorithms that require clustering step (KO, ECKO and ECDL). In the two cases, we start with a qualitative comparison of the returned results. The brain maps are ternary: all regions outside $\hat{S}$ are zeroed, while regions in $\hat{S}$ get a value of $+1$ or -1, depending on whether the contribution to the prediction is positive or negative. For ECKO, a vote is performed to decide whether a voxel is more frequently in a cluster with positive or negative weight.

4 Results

Synthetic data. A strong demonstration of how ECKO makes an improvement in stabilizing the single-clustering Knockoffs (CKO) is shown in Fig. 2. There is a clear distinction between selection of the orange area at lower right and the blue area at upper right in the CKO result, compared to the ground truth. Moreover, CKO falsely discovers some regions in the middle of the cube. By Contrast,

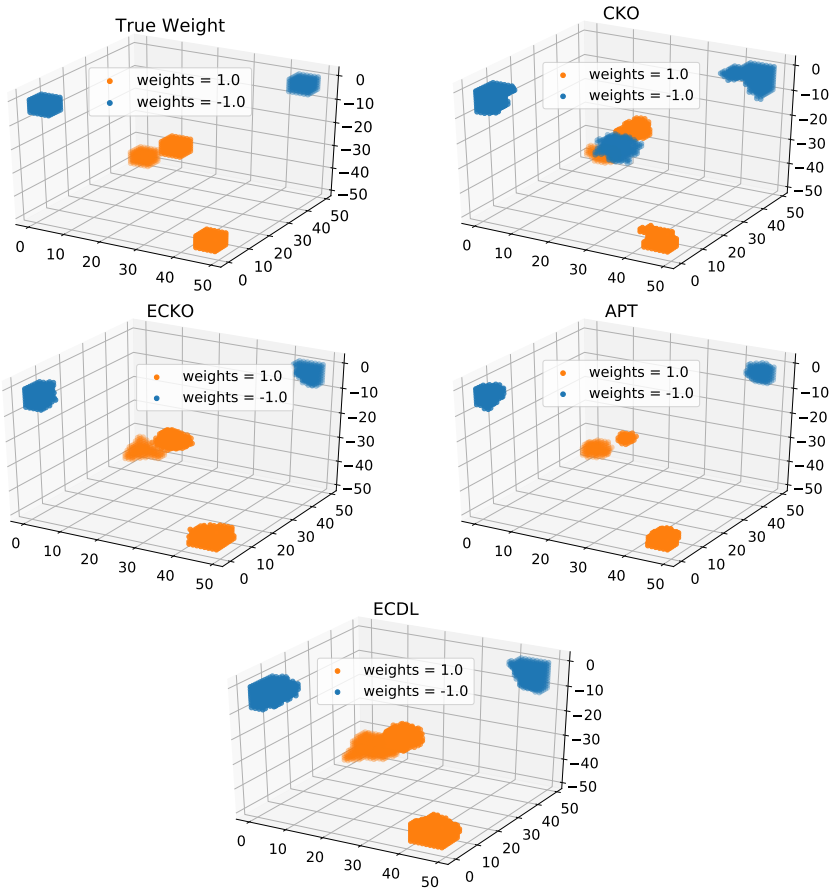
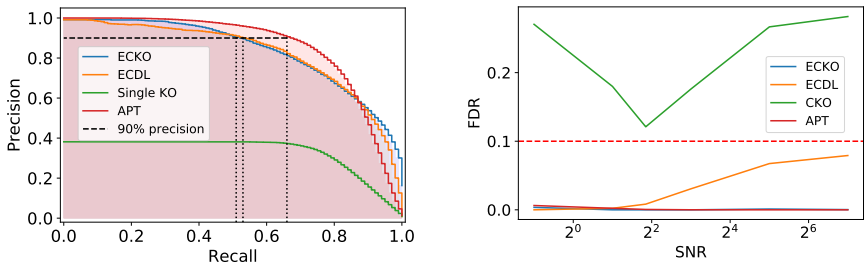


Fig. 2: Experiments on simulated data: Original 3D weight vector (top left) and inference results from CKO vs. ECKO. The single CKO run has markedly different solutions to the ground truth. Meanwhile, ECKO's solution is closer to the ground truth in the sense that altogether, it is more powerful than APT and also more precise than ECDL.

ECKO's selection is more similar to the true 3D weight cube. While it returns a wider selection than ECKO, ECDL also claims more false discoveries, most visibly in the blue area on upper-left corner. At the same time, APT returns adequate results, but is more conservative than ECKO. Fig. 3a is the result of averaging 30 simulations for the 3D brain synthetic data. ECKO and ECDL obtain almost identical precision-recall curve: for a precision of at least 90%, both methods have recall rate of around 50%. Meanwhile, CKO falls behind, and in fact it cannot reach a precision of over 40% across all recall rates. APT yields the best precision-recall compromise, slightly above ECKO and ECDL. When varying SNR (from 2^{-1} to 2^{5}) and investigating the average proportion



(a) (b)

Fig. 3: (a) Average Precision-recall curve (for $SNR \approx 3.6$) and (b) SNR-FDR curve of 30 synthetic simulations. Nominal FDR control level is 10%. ECKO shows substantially better results than CKO and is close to ECDL. APT obtains a slightly better Precision-recall curve. ECKO, ECDL and APT successfully control FDR under nominal level 0.1 where as CKO fails to.

of false discoveries (δ-FDR) made over the average of 30 simulations (Fig. 3b), we observe that CKO fails to control δ-FDR at nominal level 10% in general. Note that accurate δ-FDR control would be obtained with larger δ values, but this makes the whole procedure less useful. The ECDL controls δ-FDR at low SNR level. However, when the signal is strong, ECDL might select more false positives. ECKO, on the other hand, is always reliable —albeit conservative— keeping FDR below the nominal level even when SNR increases to larger magnitude.

Oasis & Haxby dataset. When decoding the brain signal on subject 2 of the Haxby dataset using response vector label for watching 'Face vs. House', there is a clear resemblance of selection results between ECKO and ECDL. Using an FDR threshold of 10%, both algorithms select the same area (with a difference in size), namely a face responsive region in the ventral visual cortex, and agree

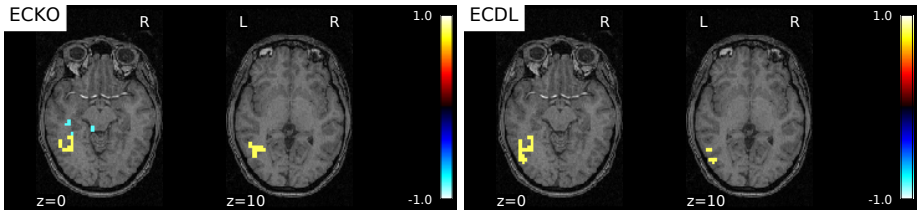


Fig. 4: Comparison of results for 2 ensembling clustered inference methods on Haxby dataset, nominal FDR=0.1. The results are similar to a large extent. No voxel region is detected by APT, therefore we omit to show the selection outcome of the method.

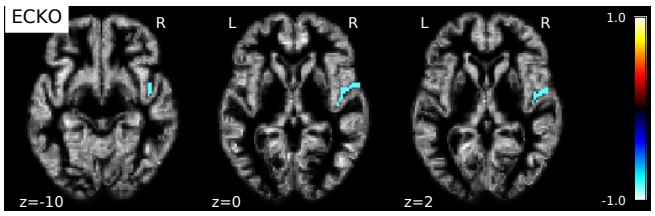


Fig. 5: Results of ECKO inference on Oasis dataset, nominal FDR=0.1. ECKO is the only method to detect significant regions. The temporal region detected by ECKO would be detected by other approaches using a less conservative threshold.

on the sign of the effect. However, on Oasis dataset, thresholding to control the FDR at 0.1 yields empty selection with ECDL and APT, while ECKO still selects some voxels. This potentially means that ECKO is statistically more powerful than ECDL and APT.

5 Conclusion

In this work, we proposed an algorithm that makes False Discovery Rate (FDR) control possible in high-dimensional statistical inference. The algorithm is an integration of clustering algorithm for dimension reduction and aggregation technique to tackle the instability of the original knockoff procedure. Evaluating the algorithm on both synthetic and brain imaging datasets shows a consistent gain of ECKO with respect to CKO in both FDR control and sensitivity. Furthermore, empirical results also suggest that the procedure achieves non-asymptotic statistical guarantees, yet requires the δ-relaxation for FDR.

The number of clusters represents a bias-variance trade-off: increasing it can reduce the bias (in fact, the value of δ), while reducing it improves the conditioning for statistical inference, hence the sensitivity of the knockoff control. We set it to 500 in our experiments. Learning it from the data is an interesting research direction.

We note that an assumption of independence between hypothesis tests is required for the algorithm to work, which is often not the case in realistic scenarios. Note that this is actually the case for all FDR-controlling procedures that rely on the BHq algorithm. As a result, making the algorithm work with relaxed assumption is a potential direction for our future study. Furthermore, the double-aggregation procedure makes the algorithm more expensive, although it results in embarrassingly parallel loops. An interesting challenge is to reduce the computation cost of this procedure. Another avenue to explore for the future is novel generative schemes for knockoff, based e.g. on deep adversarial approaches.

Acknowledgement This research is supported by French ANR (project FAST-BIG ANR-17-CE23-0011) and Labex DigiCosme (project ANR-11-LABEX-0045-

DIGICOSME). The authors would like to thank Sylvain Arlot and Matthieu Lerasle for fruitful discussions and helpful comments.

References

1. Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., Varoquaux, G.: Machine learning for neuroimaging with scikit-learn. Frontiers in Neuroinformatics 8, 14 (2014)
2. Barber, R.F., Candès, E.J.: Controlling the false discovery rate via knockoffs. The Annals of Statistics 43(5), 2055–2085 (Oct 2015), http://arxiv.org/abs/1404.5609, arXiv: 1404.5609
3. Benjamini, Y., Hochberg, Y.: Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological) 57(1), 289–300 (1995), https://www.jstor.org/stable/2346101
4. Bühlmann, P., Rütimann, P., van de Geer, S., Zhang, C.H.: Correlated variables in regression: Clustering and sparse estimation. Journal of Statistical Planning and Inference 143(11), 18351858 (2013)
5. Candès, E., Fan, Y., Janson, L., Lv, J.: Panning for gold: model-x knockoffs for high dimensional controlled variable selection. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 80(3), 551–577
6. Chevalier, J.A., Salmon, J., Thirion, B.: Statistical inference with ensemble of clustered desparsified lasso. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. pp. 638–646. Springer International Publishing, Cham (2018)
7. Gaonkar, B., Shinohara, R.T., Davatzikos, C.: Interpreting support vector machine models for multivariate group wise analysis in neuroimaging. Medical Image Analysis 24(1), 190204 (2015)
8. Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P.: Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293(5539), 2425–2430 (2001)
9. Hoyos-Idrobo, A., Varoquaux, G., Schwartz, Y., Thirion, B.: Frem scalable and stable decoding with fast regularized ensemble of models. NeuroImage 180, 160 – 172 (2018), http://www.sciencedirect.com/science/article/pii/S1053811917308182, new advances in encoding and decoding of brain signals
10. Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L.: Open access series of imaging studies (oasis): Cross-sectional mri data in young, middle aged, nondemented, and demented older adults. Journal of Cognitive Neuroscience 19(9), 14981507 (2007)
11. Meinshausen, N., Meier, L., Bhlmann, P.: p-values for high-dimensional regression. Journal of the American Statistical Association 104(488), 1671–1681 (2009)
12. Weichwald, S., Meyer, T., zdenizci, O., Schlkopf, B., Ball, T., Grosse-Wentrup, M.: Causal interpretation rules for encoding and decoding models in neuroimaging. NeuroImage 110, 48 – 59 (2015), http://www.sciencedirect.com/science/article/pii/S105381191500052X
13. Zhang, C.H., Zhang, S.S.: Confidence intervals for low dimensional parameters in high dimensional linear models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 76(1), 217242 (Mar 2014)