FDR Controlled Grouped Variable Selection for High-Dimensional Complex-Valued Data

Fabian Scheidt, Michael Muma
Robust Data Science Group
Technische Universität Darmstadt
64283, Darmstadt, Germany
{fabian.scheidt, michael.muma}@tu-darmstadt.de

***Abstract*—False discovery rate (FDR) control serves as a critical safeguard in high-dimensional statistical analysis, increasing reliability and ensuring reproducibility of discoveries. Although extensively studied for real-valued data, FDR methodologies remain underdeveloped for the complex-valued data domain despite their importance in signal processing, physics, and engineering applications. This work confronts a central challenge in FDR-controlled analysis: settings in which the data contain groups of highly correlated variables. To achieve this, we advance our *Complex-Valued Terminating-Random Experiments (CT-Rex)* framework by developing the *CT-Rex+GVS*, a grouped variable selector. The method comes in two variants: the elastic net (*EN*) and the informed elastic net (*IEN*), both with isotropic and phase agnostic regularization. We validate the framework by benchmarking it in the selection of complex-valued grouped variables for linear regression and in the estimation of the single-snapshot direction-of-arrival (DOA) based on compressed sensing using uniform linear arrays (ULA). The conducted numerical experiments confirm the FDR control property and demonstrate favorable performance compared to existing approaches.**

***Index Terms*—false discovery rate (FDR) control, CT-Rex, grouped-variable selection, single-snapshot multi-source detection, DOA estimation.**

I. Introduction

The control of the false discovery rate (FDR) is widely used to maximize the number of true positive detections while constraining the expected proportion of false positives in selection problems. Ensuring FDR control is crucial for preserving the reliability and interpretability of statistical analyses. This is particularly beneficial in high-dimensional data scenarios, where the risk of spurious detections increases due to the large number of hypotheses tested. Although substantial research has focused on developing FDR control methods for real-valued data (for low- and high-dimensions [1]–[5]), the extension to complex-valued data, a key requirement in many signal processing, engineering, and quantum physics applications, remains relatively underexplored. This work extends our previous work on this topic, i.e., the complex-valued *T-Rex* (*CT-Rex*) selector [6], and investigates the selection of variables with data-driven group structure in sparse support regimes. Thus, our work contributes to the existing literature on complex-valued statistics [7]–[12], which has primarily been focused on probability theory fundamentals, estimation methodologies, and challenges related to non-circularity. Building upon the established Terminating-Random Experiments (*T-Rex*) selector framework [4], [5], [13]–[15], we propose the extension of the *CT-Rex* selector framework by adopting the elastic net (*EN*) [16], [17] and the informed elastic net (*IEN*) [18] variants under isotropic and phase agnostic regularization with FDR control and tailoring them for the selection of grouped complex-valued variables. This culminates in the proposed complex-valued *CT-Rex+GVS* selector. Our research aims to develop and establish robust tools for FDR control in the complex domain, with potential applications in compressed-sensing-based DOA estimation [19]–[22], mechanical engineering [23]–[25], and magnetic resonance imaging [26], thus bridging an important gap in the signal processing and engineering literature.

Organization: Sec. II briefly recaps the *CT-Rex* Selector framework. Sec. III details the modifications for the elastic net (*EN*) and informed elastic net (*IEN*) theory for the complex-data case. Sec. IV provides numerical simulation results for a complex regression model with sparse grouped variable support, and additionally a single-snapshot DOA estimation scenario with extended and isolated targets. Sec. V concludes.

The first author is supported by the federal German BMBFR Clusters4Future initiative curATime, within the curAIsig project under grant number 03ZU1202MA. The second author is supported by the ERC Starting Grant ScReeningData under grant number 101042407.

II. The Complex *T-Rex* Selector

The recently proposed *CT-Rex* selector provably controls a user-defined target FDR while maximizing the number of selected variables, especially in complex-valued high-dimensional data settings [6]. It achieves this by modeling and fusion of multiple early terminated complex-valued random experiments, in which pseudo-randomly generated dummy variables compete with the original variables. Its inputs are:

1) The predictor matrix $\boldsymbol{X} = [\boldsymbol{x}_1 \dots \boldsymbol{x}_p]$, whose p predictors each contain n samples, i.e., $\boldsymbol{X} \in \mathbb{C}^{n \times p}$.
2) The response vector $\boldsymbol{y} = [y_1 \dots y_n]^\top$, $\boldsymbol{y} \in \mathbb{C}^n$.
3) The user-defined target FDR level $\alpha \in [0, 1]$.

A schematic overview of the *CT-Rex* selector framework is shown in Fig. 1 and its main steps are briefly summarized:
Step 1 (Generate Dummies): A set of $K > 1$ dummy matrices $\{\mathring{\boldsymbol{X}}_k = [\mathring{\boldsymbol{x}}_{k,1} \dots \mathring{\boldsymbol{x}}_{k,L}]\}_{k=1}^K$, each with L dummy variables, is generated from a circularly symmetric white probability distribution with finite mean and nonzero finite variance, e.g., the complex circularly white standard normal $\mathcal{CN}(0, \sigma^2\mathbf{I})$.
Step 2 (Append Complex Dummies): The original predictor matrix $\boldsymbol{X}$ is augmented K times and a set of enlarged predictor matrices $\{\widetilde{\boldsymbol{X}}_k = [\boldsymbol{X}\ \mathring{\boldsymbol{X}}_k]\}_{k=1}^K$ is formed. This establishes an auxiliary tool for FDR-controlled variable selection in which original and dummy variables compete for selection.
Step 3 (Complex-Valued Forward Variable Selection): A complex-valued forward variable selector is applied to a set of tuples $\{(\widetilde{\boldsymbol{X}}_k, \boldsymbol{y})\}_{k=1}^K$ and terminates each of these K random

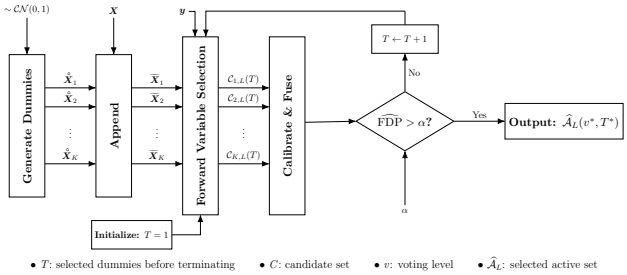


Fig. 1: Sketch of the *CT-Rex* selector.

experiments, for the first time, when $T = 1$ dummy variable enters the respective active sets. For a complex-valued linear model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{1}$$

with a complex sparse support vector $\boldsymbol{\beta} \in \mathbb{C}^p$, cardinality $\|\boldsymbol{\beta}\|_0 = s < p$, and an error term $\boldsymbol{\varepsilon} \sim \mathcal{CN}(0, \sigma^2\mathbf{I})$ we proposed the *CT-LARS* (Algorithm 1 in [6]) building upon [13], [24], [27].

Step 4 (Calibrate & Fuse): The outcome of each complex-valued random experiment is a set of candidate variables $\{\mathcal{C}_{k,L}(T)\}_{k=1}^K$ from which the T dummies are removed, and the relative occurrence of each variable is computed as:

$$\Phi_{T,L}(j) = \begin{cases} \frac{1}{K}\sum_{k=1}^K \mathbb{1}_k(j,T,L), & T \geq 1 \\ 0, & T = 0 \end{cases}. \tag{2}$$

The indicator function $\mathbb{1}_k(j,T,L)$ equals one if the jth variable was selected in the kth complex-valued random experiment. Then, a conservative estimate of the false discovery proportion (FDP), is computed and compared against the user-defined FDR threshold α. If the threshold is not exceeded, the dummy count is incremented, i.e., $T \leftarrow T+1$, and the next iteration of complex forward variable selection is computed until the FDP exceeds the target FDR α. Following this procedure, the *CT-Rex* automatically determines its parameters T, L, and v, s.t., the FDR is controlled at the user-defined target level (see Theorem 1 in [4]).

Step 5 (Output): The active set, i.e.,

$$\widehat{\mathcal{A}}_L = \{j : \Phi_{T^*,L}(j) > v^*\} \tag{3}$$

with v^* and T^*, denoting the optimal values for the voting level v and the included dummies T, which maximize the number of selected variables (see Theorem 3 in [4]).

III. THE COMPLEX *T-Rex+GVS* SELECTOR

This section introduces the modifications to the *CT-Rex*'s forward selection and dummy generation to obtain the *CT-Rex+GVS* selector. These enable the selection of correlated grouped variables for high-dimensional complex-valued data.

A. Group Variable Selection with the Elastic Net (EN) and the Informed Elastic Net (IEN)

In the complex number domain the elastic net is defined as

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}\in\mathbb{C}^p}{\arg\min} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 + \lambda_1\|\boldsymbol{\beta}\|_1 + \lambda_2\|\boldsymbol{\beta}\|_2^2 \tag{4}$$

where $\lambda_1, \lambda_2 \in \mathbb{R}_0^+$ and $\|\boldsymbol{\beta}\|_1 = \sum_{j=1}^p |\beta_j|$ and $\|\boldsymbol{\beta}\|_2^2 = \sum_{j=1}^p |\beta_j|^2$ are appropriately defined ℓ_1- and ℓ_2-norms. The grouping effect is controlled by λ_2, which is typically chosen via 10-fold cross-validation using Ridge regression based on $\boldsymbol{X}$ and $\boldsymbol{y}$. Using the *data-augmented LARS-EN* solver to solve the *EN* as a *Lasso* problem requires modifying $\boldsymbol{X}$ and $\boldsymbol{y}$ as

$$\boldsymbol{X'} = \frac{1}{\sqrt{1+\lambda_2}}\begin{pmatrix}\boldsymbol{X} \\ \sqrt{\lambda_2}\mathbf{I}_p\end{pmatrix} \quad \text{and} \quad \boldsymbol{y'} = \begin{pmatrix}\boldsymbol{y} \\ \mathbf{0}_p\end{pmatrix}, \tag{5}$$

where $\mathbf{I}_p \in \mathbb{R}^{n\times p}$ ensures the isotropic and phase-agnostic regularization approach used in this work, i.e., $\|\boldsymbol{\beta}\|_2 = \|\boldsymbol{\beta}e^{\mathrm{i}\phi}\|_2 \forall \phi \in [0, 2\pi)$. The new $\boldsymbol{X'} \in \mathbb{C}^{(n+p)\times p}$ and $\boldsymbol{y'} \in \mathbb{C}^{(n+p)}$. Future work will introduce a more efficient solver.

The informed elastic net (*IEN*) [18] incorporates information about how variables are grouped into the penalty term. It represents the mth group using a binary support vector $\mathbf{1}_m = [1_{m,1}, \ldots, 1_{m,p}]^\top \in \{0,1\}^p$ with corresponding group sizes $p_m = \sum_{j=1}^p 1_{m,j}$. The grouping itself can be known *a priori*, or obtained unsupervised, as outlined in the next section. The Lagrangian for the *IEN* is defined as

$$\mathcal{L}_{\text{IEN}}(\boldsymbol{\beta}) = \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 + \lambda_1\|\boldsymbol{\beta}\|_1 + \lambda_2\sum_{m=1}^p \frac{\|\mathbf{1}_m^\top\boldsymbol{\beta}\|_2^2}{p_m} \tag{6}$$

and its solution as

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}\in\mathbb{C}^p}{\arg\min}\, \mathcal{L}_{\text{IEN}}(\boldsymbol{\beta}). \tag{7}$$

The algorithm then refines its inputs as

$$\boldsymbol{X'} = \sqrt{\lambda_2}\begin{pmatrix}\boldsymbol{X}/\sqrt{\lambda_2} \\ \mathbf{1}_1^\top/\sqrt{p_1} \\ \ldots \\ \mathbf{1}_M^\top/\sqrt{p_M}\end{pmatrix} \quad \text{and} \quad \boldsymbol{y'} = \begin{pmatrix}\boldsymbol{y} \\ \mathbf{0}_M\end{pmatrix}. \tag{8}$$

Thus, the *IEN* appends only M rather than p entries to the data. In its augmented form, this makes it computationally more efficient than the *EN*, especially for $p \gg M$. For non-grouped variables, i.e., $M = p$ and $p_1 = \ldots = p_M = 1$, the *IEN* and *EN* are identical. Based on (5) and (8) the optimization problems can be solved as *Lasso*-type optimization problems by the *CT-LARS* as

$$\widehat{\boldsymbol{\beta}}_{\textit{CT-LARS-EN/IEN}} = \underset{\boldsymbol{\beta}\in\mathbb{C}^p}{\arg\min}\|\boldsymbol{y'} - \boldsymbol{X'}\boldsymbol{\beta}\|_2^2 + \lambda_1\|\boldsymbol{\beta}\|_1. \tag{9}$$

B. Variable Grouping and Group Dummy Generation

In the context of group variable selection, the dummy generation process of the *CT-Rex* selector must be modified, to generate K dummy matrices $\mathring{\boldsymbol{X}}_k$ that mimic the group correlation structure of $\boldsymbol{X}$. When the grouping structure is unknown, it is discovered using any hard partitioning clustering method. Our implementation uses hierarchical clustering and supports single, complete, and average linkage, as well as the Ward procedure [28]. The distance matrix $\boldsymbol{D} = 1 - |\boldsymbol{R}|$, where the correlation matrix $\mathbf{R} = \mathbb{E}[\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}]$. Groups are formed either by setting the number of clusters *a priori* or by thresholding the dendrogram at ρ_{thr}. The resulting dendrogram partitions $\boldsymbol{X}$ into M disjoint groups of variables $\{\mathcal{G}_i\}_{i=1}^M$ with $M \leq p$. For *a priori* known group structures, this step is skipped, but the dummy generation must still respect the grouping. Building upon the *T-Rex* framework (see extended calibration algorithm [4]), the *C-TRex+GVS* selector appends L dummies, being a multiple of p. Thus, a total of L/p sub-dummy matrices

are created to mimic the group-specific structure and serve as null statistics. According to the grouping sub-cluster empirical covariance matrices $\widehat{\boldsymbol{\Sigma}}_m = \frac{1}{(n-1)} \boldsymbol{X}_m^H \boldsymbol{X}_m$ are formed and the dummies for the mth group are generated from $\mathcal{CN}(\mathbf{0}, \widehat{\boldsymbol{\Sigma}}_m)$.

IV. SIMULATION STUDIES

This section presents numerical simulation results that confirm the FDR-control property and benchmark the *CT-Rex+GVS* selector. In Sec. IV-A we analyze a challenging sparse linear complex-valued regression scenario with grouped variables. In Sec. IV-B we examine a compressed sensing single-snapshot direction of arrival (DOA) estimation problem with groups representing extended targets and isolated sources.

A. Sparse Complex Regression with Group Structures

We consider a high-dimensional ($p > n$) complex sparse linear regression model, i.e., Eq. (1). The predictor matrix $\boldsymbol{X}$ is composed of M groups of variables $\mathcal{G} = \{\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_M\} \subseteq \{1, \ldots, p\}$, of non-uniform group size p_k and the Mth group has a group size of $p_M = p - \sum_{k=1}^{M-1} p_k$. The experimental parameters are set to $M = 4$ groups, $p = 500$ variables, and $n = 300$ observations. The groups $m = 1, \ldots, 3$ represent different types of structured variables, drawn from $\boldsymbol{X}_m \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_m)$ with group covariance profile

$$\boldsymbol{\Sigma}_m = \begin{cases} 1, & \ell = |i-j| = 0 \\ 0.5^{\ell} e^{\mathrm{i}\phi(\ell)}, & \ell = |i-j| \neq 0 \end{cases}, \tag{10}$$

and group phase profiles $\phi_1(\ell) = 0.2\pi\ell$, $\phi_2(\ell) = \frac{\pi}{3}\sin(2\pi 0.1\ell)$ and $\phi_3 = \phi_0 \sim \mathcal{U}[0, 2\pi) \forall \ell \neq 0$. The 4th group $\boldsymbol{X}_4 \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. For $\boldsymbol{X}$ in this simulation example, the variable indices of the structured groups are set to be $\mathcal{G}_1 = \{25, 26, 27, 28\}, \mathcal{G}_2 = \{150, 151, 152, 153, 154, 155\}, \mathcal{G}_3 = \{450, 451, 452, 453\}$, while the remaining index set $\mathcal{G}_4 = \{1, \ldots, p\} \setminus \bigcup_{m=1}^{3} G_m$ forms $\boldsymbol{X}_4$. The support of $\boldsymbol{\beta}$ is set as $\bigcup_{m=1}^{3} G_m$, i.e., $\|\boldsymbol{\beta}\|_0 = 14$. The non-zero support of $\boldsymbol{\beta} \sim \exp(\mathrm{i} \cdot \mathcal{U}(0, 2\pi))$ with $\mathrm{i} = \sqrt{-1}$, and the noise term $\boldsymbol{\varepsilon} \sim \mathcal{CN}(0, \sigma^2 \mathbf{I})$. Also, we set the target FDR $\alpha = 5\%$, and evaluate the variable selection performance for a linearly valued signal-to-noise ratio (SNR) grid $\mathrm{SNR} \in \{0.1, 0.5, 1, 2, 5, 10\}$ and average the results over $N_{\mathrm{MC}} = 200$ Monte Carlo simulations. We compare the performance of the *CT-Rex+GVS* selector with the *CT-Rex* and a *Model-X+* Knockoff method (*CKnock*) [6]. To use the *CKnock* method, as in [6], we must apply a real-value transformation with $\Re(\cdot)$ and $\Im(\cdot)$ denoting real- and imaginary parts, i.e.,

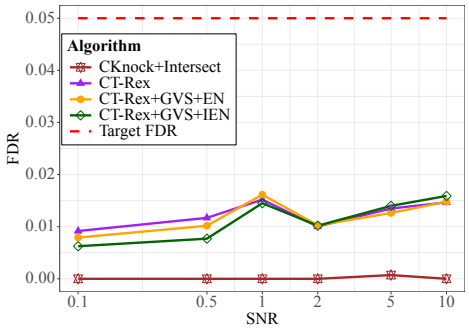
$$\boldsymbol{X}_R = \begin{pmatrix} \Re(\boldsymbol{X}) & -\Im(\boldsymbol{X}) \\ \Im(\boldsymbol{X}) & \Re(\boldsymbol{X}) \end{pmatrix}, \quad \boldsymbol{X}_R \in \mathbb{R}^{2n \times 2p}, \tag{11}$$

$$\boldsymbol{y}_R = [\Re(\boldsymbol{y}) \quad \Im(\boldsymbol{y})]^{\top}, \quad \boldsymbol{y}_R \in \mathbb{R}^{2n}, \tag{12}$$

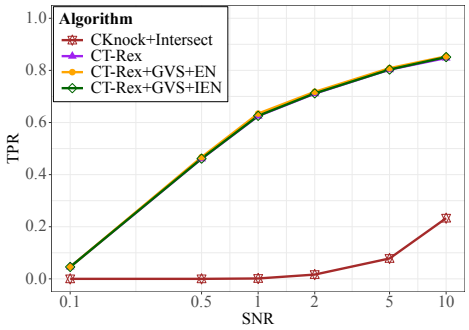
and consequently $\boldsymbol{\beta}_R \in \mathbb{R}^{2p}$. The transformation preserves the properties of complex-valued multiplication and conjugation. After determining the set of active indices, the estimated support set of $\boldsymbol{\beta}$ is split in half, accounting for real and imaginary parts, respectively. To maintain FDR control of the *CKnock* for complex-valued data, the final estimate of the support of $\boldsymbol{\beta}$ is obtained by forming the intersection, i.e.,

$$\mathrm{support}(\widehat{\boldsymbol{\beta}}) = \mathrm{support}(\Re(\widehat{\boldsymbol{\beta}})) \cap \mathrm{support}(\Im(\widehat{\boldsymbol{\beta}})). \tag{13}$$

Note that the union does not provide FDR control [6]. For the *CKnock* selector, we used Gaussian Knockoffs and an approximate semi-definite programming solver due to computational constraints, as both second-order Knockoffs and exact semi-definite programming were infeasible within the 24-hour processing limit. As shown in Fig. 2a, the proposed *CT-Rex+GVS* and its two competitors the *CT-Rex*, and the *CKnock* selector, indeed control the target FDR $\alpha = 5\%$ for each of the SNR levels. The *CT-Rex+GVS* in its *EN* and *IEN* variants perform similar, and all three clearly outperform the *model-X+* based *CKnock* methods in terms of TPR.



(a) FDR-control property (Target FDR $\alpha = 5\%$).



(b) TPR performance.

Fig. 2: Complex linear regression with FDR control for grouped variables and varying SNR levels.

B. Single-Snapshot Compressed Beamforming with FDR Control for DOA Estimation of Extended and Isolated Targets

The following scenario concerns the joint detection and estimation of the direction of arrival (DOA) of multiple source signals. This is fundamentally important in applications such as radar and sonar systems. The work of [19] presented compressed sensing (CS) algorithms for DOA estimation, showing that the array output $\boldsymbol{y}$ can be modeled as a sparse linear combination of steering vectors in a discretized DOA space. Given a support estimate, the DOA estimates are obtained by mapping them on the grid. Note that the performance of compressed beamforming algorithms [19] depends on the grid resolution. A denser grid increases mutual coherence among the basis vectors, which leads to poorer recovery performance. Inspired by [20]–[22], we adopt a *Lasso*-type sparse regression approach and compare our method to their proposed algorithms: the *adaptive Lasso (ALASSO)*, the *adaptive Elastic Net (AEN)*, and the *Sequential Adaptive Elastic Net (SAEN)*.

The sensor array is a uniform linear array (ULA) of M sensors aiming at DOA estimation for $\theta \in [-90°, 90°)$ relative to the array axis. We assume narrowband processing, and make the far-field assumption (i.e., propagation radius $\gg$ array size), resulting in a plane wave signal model. The steering vector of a ULA with half-wavelength inter-element spacing for a source at angle θ is given by

$$\mathbf{a}(\theta) = \frac{1}{\sqrt{M}} \left[1 \; e^{\mathrm{i}\pi \sin(\theta)} \; \dots \; e^{\mathrm{i}\pi(M-1)\sin(\theta)}\right]^\top. \tag{14}$$

For $Q < M$ sources with distinct DOAs $\{\theta_q\}_{q=1}^{Q}$, a single-snapshot ULA measurement is modeled as

$$\boldsymbol{y} = \mathbf{A}(\theta)\tilde{\mathbf{s}} + \boldsymbol{\varepsilon}\,, \tag{15}$$

where $\mathbf{A}(\theta) = [\mathbf{a}(\theta_1) \dots \mathbf{a}(\theta_Q)]$ is the steering matrix for the true, continuous-valued source directions θ_q, and $\tilde{\mathbf{s}} \in \mathbb{C}^Q$ is the source signal vector. To leverage CS methods, the continuous angular domain is discretized into a uniform grid $\{\theta_g\}_{g=1}^{G}$, leading to the CS model

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\,, \tag{16}$$

where $\boldsymbol{\Phi} \in \mathbb{C}^{M\times G}$ is the dictionary matrix whose columns are the steering vectors $\mathbf{a}(\theta_g)$ for each grid point, and $\boldsymbol{\beta} \in \mathbb{C}^G$ being sparse and of cardinality $\|\boldsymbol{\beta}\|_0 = Q$ with nonzero entries corresponding to source locations on the grid. This formulation enables sparse recovery since typically $Q \ll G$. However, model mismatch occurs if true DOAs lie off-grid, potentially degrading estimation accuracy.

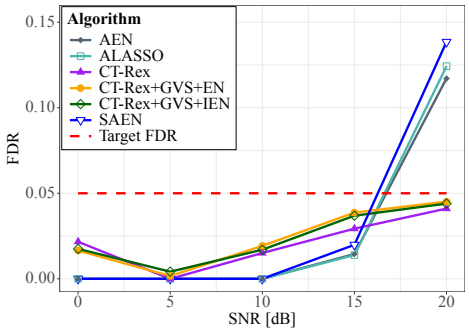
We consider a ULA with $M = 64$ sensor elements and an angular grid resolution of $1°$. Furthermore, we assume $Q = 8$ heterogeneous sources originating from two extended targets, represented by signal groups of support indices $\mathcal{G}_1$ and $\mathcal{G}_2$, and two isolated targets within group $\mathcal{G}_3$, i.e., $\mathcal{G}_1 = \{\theta : \theta \in \{-45°, -44°, -43°\}\}, \mathcal{G}_2 = \{\theta : \theta \in \{21°, 22°, 23°\}\}, \mathcal{G}_3 = \{\theta : \theta \in \{-60°, 50°\}\}$. The source signals for the two extended targets were designed with group-specific parameters. For group $\mathcal{G}_k (k \in 1, 2)$, the amplitudes $\boldsymbol{\alpha}_k \sim \mathcal{N}(\boldsymbol{\mu}_{\alpha_k}, \sigma^2_{\alpha_k}\boldsymbol{C}_k)$ and phases $\boldsymbol{\phi}_k \sim \mathcal{N}(\boldsymbol{\mu}_{\phi_k}, \sigma^2_{\phi_k}\boldsymbol{C}_k)$ $\bmod 2\pi$, where $\boldsymbol{C}_k$ is the within-group correlation matrix

$$\mathbf{C}_k = \mathbf{I} + \rho_k \mathbf{1}\mathbf{1}^\top, \tag{17}$$

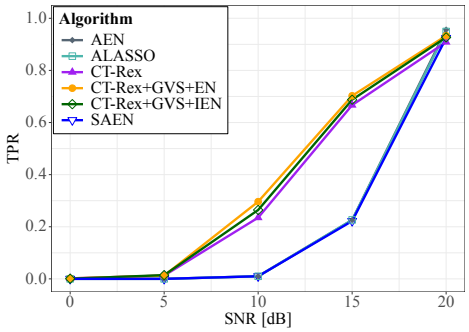
and $\rho_k \in [0, 1)$ as intra-group correlation coefficient. The final signals for group $\mathcal{G}_k$ are computed as $\boldsymbol{s}_k = \boldsymbol{\alpha}_k \odot e^{\mathrm{i}\boldsymbol{\phi}_k}$ where $\odot$ denotes element-wise multiplication. The two isolated sources of $\mathcal{G}_3$ were modeled as $\boldsymbol{s}_{\text{iso}} = e^{\mathrm{i}\mathcal{U}(0,2\pi)}$. The comparison methods - *ALASSO*, *AEN*, and *SAEN* - use the generalized information criterion, according to [22], to estimate the source number. In contrast, the *CT-Rex+GVS* and the *CT-Rex* selector do it blind. The *CT-Rex+GVS* selector used the Ward procedure to uniformly cluster the steering matrix into angular sectors. For the simulation, we conducted 200 Monte Carlo trials per algorithm. In each trial noise was generated as $\boldsymbol{\varepsilon} \sim \mathcal{CN}(0, \sigma^2\mathbf{I})$, and the performance was evaluated across the SNR grid $\{0, 5, 10, 15, 20\}$ dB at target FDR $\alpha = 5\%$. In Figs. 3a-3b we report FDR and TPR performance. The results show that the FDR is controlled in this scenario by *CT-Rex+GVS* and *CT-Rex* across all SNR levels. Notably, the TPR performance of both exceeds that of the competing methods particularly at low to mid-range SNR. At 20 dB, the competitors clearly exceed the target FDR.

TABLE I: Parametrization for $\mathcal{G}_k \in \{1, 2\}$.

Group	$\boldsymbol{\mu}_{\alpha_k}$	σ_{α_k}	$\boldsymbol{\mu}_{\phi_k}$	σ_{ϕ_k}	ρ_k
$\mathcal{G}_1$	1	0.05	$\pi/4$	0.1	0.8
$\mathcal{G}_2$	0.8	0.02	$\pi/3$	0.05	0.9



(a) FDR performance.



(b) TPR performance.

Fig. 3: Compressed sensing based grouped DOA estimation for varying SNR levels with FDR control.

V. CONCLUSIONS

This work introduced the *CT-Rex+GVS* selector, extending the *CT-Rex* framework to enable grouped variable selection with FDR control in high-dimensional complex-valued data. By integrating elastic- and informed elastic net regularization, the method achieves robust FDR control while improving selection accuracy. Simulations in grouped sparse regression and DOA estimation demonstrate the method's ability to control the FDR. Additionally, we observe a significant increase in TPR compared to competing FDR-controlling methods. These results highlight its potential for variable selection in complex-domain applications, addressing a critical gap in signal processing and engineering. Future work will develop efficient pathwise LARS-EN solvers for large scale complex-valued problems, investigate data-adaptive grouping with phase-awareness, and extend to structured non-isotropic regularization.

REFERENCES

[1] Y. Benjamini and Y. Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 57, no. 1, pp. 289–300, 1995.

[2] Y. Benjamini and D. Yekutieli, "The Control of the False Discovery Rate in Multiple Testing under Dependency," *The Annals of Statistics*, vol. 29, no. 4, pp. 1165–1188, 2001.

[3] E. Candès, Y. Fan, L. Janson, and J. Lv, "Panning for Gold: Model-X Knockoffs for High-Dimensional Controlled Variable Selection," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 80, no. 3, pp. 551–577, 2018.

[4] J. Machkour, M. Muma, and D. P. Palomar, "The terminating-random experiments selector: Fast high-dimensional variable selection with false discovery rate control," *Signal Processing*, vol. 231, p. 109 894, 2025.

[5] J. Machkour, M. Muma, and D. P. Palomar, "High-dimensional false discovery rate control for dependent variables," *Signal Processing*, vol. 234, p. 109 990, 2025.

[6] F. Scheidt, J. Machkour, and M. Muma, "FDR Control for Complex-Valued Data with Application in Single Snapshot Multi-Source Detection and DOA Estimation," in *Proc. 50th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2025, pp. 1–5.

[7] T. Adali, P. J. Schreier, and L. L. Scharf, "Complex-Valued Signal Processing: The Proper Way to Deal With Impropriety," *IEEE Trans. Signal. Process.*, vol. 59, no. 11, pp. 5101–5125, 2011.

[8] A. M. Zoubir, V. Koivunen, E. Ollila, and M. Muma, *Robust Statistics for Signal Processing*. Cambridge University Press, 2018.

[9] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*. Cambridge University Press, 2010.

[10] B. Picinbono, "On Circularity," *IEEE Trans. Signal. Process.*, vol. 42, no. 12, pp. 3473–3482, 1994.

[11] J. Eriksson, E. Ollila, and V. Koivunen, "Essential Statistics and Tools for Complex Random Variables," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5400–5408, 2010.

[12] A. van den Bos, "The Multivariate Complex Normal Distribution—A Generalization," *IEEE Transactions on Information Theory*, vol. 41, no. 2, pp. 537–539, 1995.

[13] J. Machkour, S. Tien, D. P. Palomar, and M. Muma, *tlars: The T-LARS Algorithm: Early-Terminated Forward Variable Selection*, R package version 1.0.1, 2024. [Online]. Available: https://CRAN.R-project.org/package=tlars.

[14] J. Machkour, S. Tien, D. P. Palomar, and M. Muma, *TRexSelector: T-Rex Selector: High-Dimensional Variable Selection and FDR Control*, R package version 1.0.0, 2024. [Online]. Available: https://CRAN.R-project.org/package=TRexSelector.

[15] F. Scheidt, J. Machkour, and M. Muma, "Solving FDR-Controlled Sparse Regression Problems with Five Million Variables on a Laptop," in *Proc. 9th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 2023, pp. 116–120.

[16] H. Zou and T. Hastie, "Regularization and Variable Selection via the Elastic Net," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 67, no. 2, pp. 301–320, 2005.

[17] J. Machkour, M. Muma, and D. P. Palomar, "False Discovery Rate Control for Grouped Variable Selection in High-Dimensional Linear Models Using the T-Knock Filter," in *Proc. 30th European Signal Processing Conference (EUSIPCO)*, 2022, pp. 892–896.

[18] J. Machkour, M. Muma, and D. P. Palomar, "The Informed Elastic Net for Fast Grouped Variable Selection and FDR Control in Genomics Research," in *Proc. 9th IEEE International Workshop on Computational Advanced in Multi-Sensor Processing (CAMSAP)*, 2023, pp. 466–470.

[19] D. Malioutov, M. Cetin, and A. S. Willsky, "A Sparse Signal Reconstruction Perspective for Source Localization with Sensor Arrays," *IEEE Trans. Signal. Process.*, vol. 53, no. 8, pp. 3010–3022, 2005.

[20] M. N. Tabassum and E. Ollila, "Single-Snapshot DoA Estimation using Adaptive Elastic Net in the Complex Domain," in *Proc. 4th International Workshop on Compressed Sensing Theory and its Applications to Radar, Sonar and Remote Sensing (CoSeRa)*, 2016, pp. 197–201.

[21] M. N. Tabassum and E. Ollila, "Sequential Adaptive Elastic Net Approach for Single-Snapshot Source Localization," *The Journal of the Acoustical Society of America*, vol. 143, no. 6, pp. 3873–3882, 2018.

[22] M. N. Tabassum and E. Ollila, "Simultaneous Signal Subspace Rank and Model Selection with an Application to Single-Snapshot Source Localization," in *Proc. 26th European Signal Processing Conference (EUSIPCO)*, 2018, pp. 1592–1596.

[23] S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Discovering Governing Equations from Data by Sparse Identification of Nonlinear Dynamical Systems," *Proceedings of the National Academy of Sciences*, vol. 113, no. 15, pp. 3932–3937, 2016.

[24] J. Graff, M. J. Ringuette, T. Singh, and F. D. Lagor, "Reduced-Order Modeling for Dynamic Mode Decomposition without an Arbitrary Sparsity Parameter," *AIAA Journal*, vol. 58, no. 9, pp. 3919–3931, 2020.

[25] M. R. Jovanović, P. J. Schmid, and J. W. Nichols, "Sparsity-Promoting Dynamic Mode Decomposition," *Physics of Fluids*, vol. 26, no. 2, 2014.

[26] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The Application of Compressed Sensing for Rapid MR Imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

[27] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least Angle Regression," *The Annals of Statistics*, vol. 32, no. 2, pp. 407–499, 2004.

[28] F. Murtagh and P. Contreras, "Algorithms for hierarchical clustering: An overview," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 2, no. 1, pp. 86–97, 2012.