# Signal Decoding in Multicorrelated High-Dimensional Spaces

Roman Isachenko

Skoltech advisor: Maxim Fedorov
MIPT advisor: Vadim Strijov

April 27, 2018.

# Signal decoding problem

## Goal

- Investigate input, latent, and target spaces for signal decoding problem.
- Build a stable model for time series decoding in the case of multicorrelated object description.
- Suggest dimensionality reduction algorithm for signal decoding problem.

## Challenge

- Measurements of spatial-temporal data are interacted since sensors are close to each other.
- Target variable is a vector whose elements are dependent.

## Solution

Propose to use joint description of independent and target variables. This description allows to reduce the multicorrelation and to build stable adequate model with acceptable accuracy.

# Related works


- Katrutsa A., Strijov V. Comprehensive study of feature selection methods to solve multicollinearity problem according to evaluation criteria. // *Expert Systems with Applications* 76, 2017.

- Li J. et al. Feature selection: A data perspective //*ACM Computing Surveys (CSUR)* 50(6), 2017.

- Eliseyev A. et al. Iterative N-way partial least squares for a binary self-paced brain–computer interface in freely moving animals //*Journal of neural engineering* 4(8), 2011.

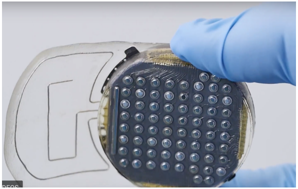
- Rodriguez-Lujan I. et al. Quadratic programming feature selection // *Journal of Machine Learning Research* 11(Apr), 2010.

- Motrenko A., Strijov V. Multi-way Feature Selection for ECoG-based Brain-Computer Interface // *Expert Systems with Applications* Submitted to the journal.

# Application: Brain Computer Interface (BCI)

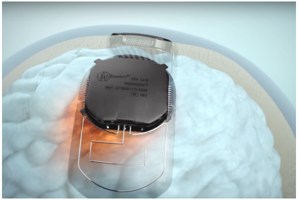
## Aim

Develop systems to help people with a severe motor control disability recover mobility.



## Hypothesis

"When we imagine making a movement, we trigger the same electrical activity in the motor cortex of the brain as when we actually perform that activity."*



## Solution

Record electrical signals – electrocorticograms (ECoG), decode them to drive complex objects, for example, to move the limbs of an exoskeleton.

* `http://clinatec.fr`;

---

Eliseyev A. et al. CLINATEC BCI platform based on the ECoG-recording implant WIMAGINE and the innovative signal-processing: preclinical results, 2014.

# Multivariate regression

## Model

Forecast a dependent variable $\mathbf{y} \in \mathbb{R}^r$ from an independent input object $\mathbf{x} \in \mathbb{R}^n$

$$\mathbf{y} = \mathbf{\Theta}\mathbf{x} + \boldsymbol{\varepsilon}, \quad \mathbf{\Theta} \in \mathbb{R}^{r \times n}$$

## Given

Dataset $(\mathbf{X}, \mathbf{Y})$, design matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$, target matrix $\mathbf{Y} \in \mathbb{R}^{m \times r}$,

$$\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_m]^T = [\boldsymbol{\chi}_1, \dots, \boldsymbol{\chi}_n]; \quad \mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_m]^T = [\boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_r].$$

## Error function

$$S(\mathbf{\Theta}|\mathbf{X}, \mathbf{Y}) = \left\| \underset{m \times r}{\mathbf{Y}} - \underset{m \times n}{\mathbf{X}} \cdot \underset{r \times n}{\mathbf{\Theta}}^T \right\|_2^2 \rightarrow \min_{\mathbf{\Theta}}.$$

$$\mathbf{\Theta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}.$$

The linear dependent columns of the matrix $\mathbf{X}$ leads to an instable solution. To avoid the strong linear dependence, feature selection and dimensionality reduction techniques are used.

# Feature selection problem statement

## Goal

Find the index set $\mathcal{A} = \{1, \dots, n\}$ of $\mathbf{X}$ columns.

## Quality Criteria

To select the set $\mathcal{A}$ among all possible $2^n - 1$ subsets, introduce the feature selection quality criteria

$$\mathcal{A} = \underset{\mathcal{A}' \subseteq \{1, \dots, n\}}{\arg\max}\; Q(\mathcal{A}' | \mathbf{X}, \mathbf{Y}).$$

Once the solution $\mathcal{A}$ is known:

$$S(\mathbf{\Theta}_{\mathcal{A}} | \mathbf{X}_{\mathcal{A}}, \mathbf{Y}) = \left\| \mathbf{Y} - \mathbf{X}_{\mathcal{A}} \mathbf{\Theta}_{\mathcal{A}}^{T} \right\|_2^2 \to \min_{\mathbf{\Theta}_{\mathcal{A}}},$$

where the subscript $\mathcal{A}$ indicates columns with indices from the set $\mathcal{A}$.

$$\|\boldsymbol{\nu} - \mathbf{X}\boldsymbol{\theta}\|_2^2 \to \min_{\boldsymbol{\theta} \in \mathbb{R}^n}.$$

## Quadratic Programming Feature Selection

$$(1-\alpha) \cdot \underbrace{\mathbf{a}^T\mathbf{Q}\mathbf{a}}_{\text{Sim}(\mathbf{X})} - \alpha \cdot \underbrace{\mathbf{b}^T\mathbf{a}}_{\text{Rel}(\mathbf{X}, \boldsymbol{\nu})} \to \min_{\substack{\mathbf{a} \geq \mathbf{0}_n \\ \mathbf{1}_n^T\mathbf{a}=1}}.$$

- $\mathbf{a} \in \mathbb{R}^n$ — feature importances;
- $\mathbf{Q} \in \mathbb{R}^{n \times n}$ - pairwise feature similarities;
- $\mathbf{b} \in \mathbb{R}^n$ - feature relevances to the target vector.

$$j \in \mathcal{A}^* \Leftrightarrow a_j > \tau$$

## Quality Criteria

$$\mathcal{A} = \operatorname*{arg\,max}_{\mathcal{A}' \subseteq \{1,\dots,n\}} Q(\mathcal{A}' | \mathbf{X}, \boldsymbol{\nu}) \Leftrightarrow \operatorname*{arg\,min}_{\mathbf{a} \geq \mathbf{0}_n,\, \mathbf{1}_n^T\mathbf{a}=1} \left[\mathbf{a}^T\mathbf{Q}\mathbf{a} - \alpha \cdot \mathbf{b}^T\mathbf{a}\right].$$

## Quality Criteria

$$\mathcal{A} = \mathop{\arg\max}_{\mathcal{A}' \subseteq \{1,\dots,n\}} Q(\mathcal{A}' | \mathbf{X}, \boldsymbol{\nu}) \Leftrightarrow \mathop{\arg\min}_{\mathbf{a} \geq \mathbf{0}_n,\, \mathbf{1}_n^T \mathbf{a} = 1} \left[\mathbf{a}^T \mathbf{Q} \mathbf{a} - \alpha \cdot \mathbf{b}^T \mathbf{a}\right].$$

## Similarity measure

- Correlation

$$|\mathrm{corr}(\mathbf{x}, \mathbf{y})| = \left| \frac{\mathrm{cov}(\mathbf{x}, \mathbf{y})}{\sqrt{\mathrm{Var}(\mathbf{x})\mathrm{Var}(\mathbf{y})}} \right|$$

- Mutual information

$$I(\mathbf{x}, \mathbf{y}) = \int \int p(\mathbf{x}, \mathbf{y}) \log \frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{x}) p(\mathbf{y})} d\mathbf{x} d\mathbf{y}.$$

$$\mathbf{Q} = \left\{ \left| \mathrm{corr}(\chi_i, \chi_j) \right| \right\}_{i,j=1}^{n}, \quad \mathbf{b} = \left\{ \left| \mathrm{corr}(\chi_i, \boldsymbol{\nu}) \right| \right\}_{i=1}^{n}.$$

## Statement

In the case of semidefinite matrix $\mathbf{Q}$ the QPFS problem is convex.

$$\mathbf{Q} \to \mathbf{Q} - \lambda_{\min} \mathbf{I}$$

## Multivariate QPFS

### Relevance aggregation

$$\mathbf{Q} = \left\{ \left|\mathrm{corr}(\chi_i, \chi_j)\right| \right\}_{i,j=1}^{n}, \quad \mathbf{b} = \left\{ \sum_{k=1}^{r} \left|\mathrm{corr}(\chi_i, \boldsymbol{\nu}_k)\right| \right\}_{i=1}^{n}.$$

This approach does not use the dependencies in the columns of the matrix $\mathbf{Y}$.

### Example:

$$\mathbf{X} = [\chi_1, \chi_2, \chi_3], \quad \mathbf{Y} = [\underbrace{\boldsymbol{\nu}_1, \boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_1}_{r-1}, \boldsymbol{\nu}_2],$$

$$\mathbf{Q} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0.8 \\ 0 & 0.8 & 1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \underbrace{\begin{matrix} 0.4 & \dots & 0.4 \\ 0.5 & \dots & 0.5 \\ 0.8 & \dots & 0.8 \end{matrix}}_{r-1} & \begin{matrix} 0 \\ 0.8 \\ 0.1 \end{matrix} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} (r-1)\cdot 0.4 + 0 \\ (r-1)\cdot 0.5 + 0.8 \\ (r-1)\cdot 0.8 + 0.1 \end{bmatrix}$$

Best subset: $[\chi_1, \chi_2]$.

QPFS ($r = 2$): $\mathbf{a} = [\mathbf{0.37}, \mathbf{0.61}, 0.02]$.
QPFS ($r = 5$): $\mathbf{a} = [\mathbf{0.40}, 0.17, \mathbf{0.43}]$.

Penalize correlated targets by Sim($\mathbf{Y}$)

$$\alpha_1 \cdot \underbrace{\mathbf{a}_x^T \mathbf{Q}_x \mathbf{a}_x}_{\text{Sim}(\mathbf{X})} - \alpha_2 \cdot \underbrace{\mathbf{a}_x^T \mathbf{B} \mathbf{a}_y}_{\text{Rel}(\mathbf{X},\mathbf{Y})} + \alpha_3 \cdot \underbrace{\mathbf{a}_y^T \mathbf{Q}_y \mathbf{a}_y}_{\text{Sim}(\mathbf{Y})} \to \min_{\substack{\mathbf{a}_x \geq \mathbf{0}_n,\, \mathbf{1}_n^T \mathbf{a}_x = 1 \\ \mathbf{a}_y \geq \mathbf{0}_r,\, \mathbf{1}_r^T \mathbf{a}_y = 1}} .$$

$$\mathbf{Q}_x = \left\{ \left| \text{corr}(\boldsymbol{\chi}_i, \boldsymbol{\chi}_j) \right| \right\}_{i,j=1}^{n}, \ \mathbf{Q}_y = \left\{ \left| \text{corr}(\boldsymbol{\nu}_i, \boldsymbol{\nu}_j) \right| \right\}_{i,j=1}^{r}, \ \mathbf{B} = \left\{ \left| \text{corr}(\boldsymbol{\chi}_i, \boldsymbol{\nu}_j) \right| \right\}_{\substack{i=1,\ldots,n \\ j=1,\ldots,r}} .$$

$$\alpha_1 + \alpha_2 + \alpha_3 = 1 \quad \alpha_i \geq 0,\ i = 1, 2, 3.$$

### Statement

*Balance between the terms Sim($\mathbf{X}$) and Rel($\mathbf{X}, \mathbf{Y}$) for the problem is achieved by the following coefficients:*

$$\alpha_1 = \frac{(1-\alpha_3)\overline{\mathbf{B}}}{\overline{\mathbf{Q}}_x + \overline{\mathbf{B}}}; \quad \alpha_2 = \frac{(1-\alpha_3)\overline{\mathbf{Q}}_x}{\overline{\mathbf{Q}}_x + \overline{\mathbf{B}}}; \quad \alpha_3 \in [0, 1],$$

*where $\overline{\mathbf{Q}}_x$, $\overline{\mathbf{B}}$ are the mean values of $\mathbf{Q}_x$ and $\mathbf{B}$ respectively.*

## Multivariate QPFS

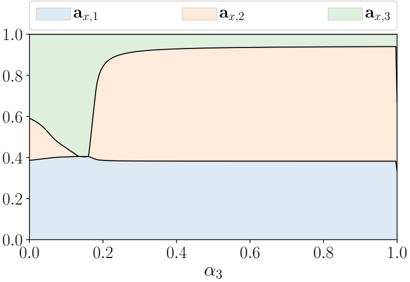
Example:

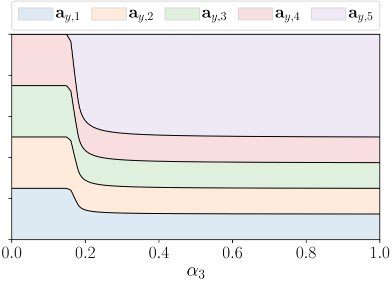
$$\mathbf{X} = [\chi_1, \chi_2, \chi_3], \quad \mathbf{Y} = [\underbrace{\nu_1, \nu_1, \ldots, \nu_1}_{r-1}, \nu_2],$$

$$\mathbf{Q} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0.8 \\ 0 & 0.8 & 1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \underbrace{\begin{matrix} 0.4 & \ldots & 0.4 \\ 0.5 & \ldots & 0.5 \\ 0.8 & \ldots & 0.8 \end{matrix}}_{r-1} & \begin{matrix} 0 \\ 0.8 \\ 0.1 \end{matrix} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} (r-1)\cdot 0.4 + 0 \\ (r-1)\cdot 0.5 + 0.8 \\ (r-1)\cdot 0.8 + 0.1 \end{bmatrix}$$

$$\mathbf{Q}_x = \mathbf{Q};\ \mathbf{Q}_y : \text{corr}(\nu_1, \nu_2) = 0.2 \text{ all others entries } = 1.$$

Best subset: $[\chi_1, \chi_2]$.



## Min-max / Max-min

Symmetric Importances penalizes targets that are correlated and are not sufficiently explained by the features.

$$\alpha_1 \cdot \underbrace{\mathbf{a}_x^T \mathbf{Q}_x \mathbf{a}_x}_{\mathrm{Sim}(\mathbf{X})} - \alpha_2 \cdot \underbrace{\mathbf{a}_x^T \mathbf{B} \mathbf{a}_y}_{\mathrm{Rel}(\mathbf{X},\mathbf{Y})} \to \min_{\substack{\mathbf{a}_x \geq \mathbf{0}_n, \\ \mathbf{1}_n^T \mathbf{a}_x = 1}} ; \quad \alpha_3 \cdot \underbrace{\mathbf{a}_y^T \mathbf{Q}_y \mathbf{a}_y}_{\mathrm{Sim}(\mathbf{Y})} + \alpha_2 \cdot \underbrace{\mathbf{a}_x^T \mathbf{B} \mathbf{a}_y}_{\mathrm{Rel}(\mathbf{X},\mathbf{Y})} \to \min_{\substack{\mathbf{a}_y \geq \mathbf{0}_r, \\ \mathbf{1}_r^T \mathbf{a}_y = 1}} .$$

### Min-max / Max-min

$$\min_{\substack{\mathbf{a}_x \geq \mathbf{0}_n \\ \mathbf{1}_n^T \mathbf{a}_x = 1}} \max_{\substack{\mathbf{a}_y \geq \mathbf{0}_r \\ \mathbf{1}_r^T \mathbf{a}_y = 1}} \left( \text{or} \max_{\substack{\mathbf{a}_y \geq \mathbf{0}_r \\ \mathbf{1}_r^T \mathbf{a}_y = 1}} \min_{\substack{\mathbf{a}_x \geq \mathbf{0}_n \\ \mathbf{1}_n^T \mathbf{a}_x = 1}} \right) \left[ \alpha_1 \cdot \underbrace{\mathbf{a}_x^T \mathbf{Q}_x \mathbf{a}_x}_{\mathrm{Sim}(\mathbf{X})} - \alpha_2 \cdot \underbrace{\mathbf{a}_x^T \mathbf{B} \mathbf{a}_y}_{\mathrm{Rel}(\mathbf{X},\mathbf{Y})} - \alpha_3 \cdot \underbrace{\mathbf{a}_y^T \mathbf{Q}_y \mathbf{a}_y}_{\mathrm{Sim}(\mathbf{Y})} \right] .$$

$$\mathcal{A} = \operatorname*{arg\,max}_{\mathcal{A}' \subseteq \{1,\dots,n\}} Q(\mathcal{A}' | \mathbf{X}, \mathbf{Y}) \Leftrightarrow \operatorname*{arg\,min}_{\mathbf{a}_x \geq \mathbf{0}_n,\, \mathbf{1}_n^T \mathbf{a}_x = 1} \left[ \max_{\substack{\mathbf{a}_y \geq \mathbf{0}_r \\ \mathbf{1}_r^T \mathbf{a}_y = 1}} f(\mathbf{a}_x, \mathbf{a}_y) \right] .$$

### Theorem

*For positive definite matrices $\mathbf{Q}_x$ and $\mathbf{Q}_y$ the max-min and min-max problems have the same optimal value.*

# Min-max / Max-min

## Lagrangian for fixed $\mathbf{a}_x$

$$L(\mathbf{a}_x, \mathbf{a}_y, \lambda, \boldsymbol{\mu}) = \alpha_1 \cdot \mathbf{a}_x^T \mathbf{Q}_x \mathbf{a}_x - \alpha_2 \cdot \mathbf{a}_x^T \mathbf{B} \mathbf{a}_y - \alpha_3 \cdot \mathbf{a}_y^T \mathbf{Q}_y \mathbf{a}_y + \lambda \cdot (\mathbf{1}_r^T \mathbf{a}_y - 1) + \boldsymbol{\mu}^T \mathbf{a}_y.$$

## Theorem

Min-max problem is equivalent to the following quadratic problem with $n + r + 1$ variables

$$\min_{\substack{\mathbf{a}_x \geq \mathbf{0}_n,\, \mathbf{1}_n^T \mathbf{a}_x = 1 \\ \lambda,\, \boldsymbol{\mu} \geq \mathbf{0}_r}} g(\mathbf{a}_y, \lambda, \boldsymbol{\mu}), \quad \text{where}$$

$$g(\mathbf{a}_x, \lambda, \boldsymbol{\mu}) = \max_{\mathbf{a}_y \in \mathbb{R}^r} L(\mathbf{a}_x, \mathbf{a}_y, \lambda, \boldsymbol{\mu}) = \mathbf{a}_x^T \left( -\frac{\alpha_2^2}{4\alpha_3} \mathbf{B} \mathbf{Q}_y^{-1} \mathbf{B}^T - \alpha_1 \cdot \mathbf{Q}_x \right) \mathbf{a}_x$$
$$- \frac{1}{4\alpha_3} \lambda^2 \cdot \mathbf{1}_r^T \mathbf{Q}_y^{-1} \mathbf{1}_r - \frac{1}{4\alpha_3} \boldsymbol{\mu}^T \mathbf{Q}_y^{-1} \boldsymbol{\mu} + \frac{\alpha_2}{2\alpha_3} \lambda \cdot \mathbf{1}_r^T \mathbf{Q}_y^{-1} \mathbf{B}^T \mathbf{a}_x$$
$$- \frac{1}{2\alpha_3} \lambda \cdot \mathbf{1}_r^T \mathbf{Q}_y^{-1} \boldsymbol{\mu} + \frac{\alpha_2}{2\alpha_3} \boldsymbol{\mu}^T \mathbf{Q}_y^{-1} \mathbf{B}^T \mathbf{a}_x + \lambda.$$

The problem is not convex. If we shift the spectrum for the matrix of quadratic form, the optimality is lost.

Drop the term $\text{Sim}(\mathbf{Y})$.

$$\min_{\substack{\mathbf{a}_x \geq \mathbf{0}_n \\ \mathbf{1}_n^T \mathbf{a}_x = 1}} \max_{\substack{\mathbf{a}_y \geq \mathbf{0}_r \\ \mathbf{1}_r^T \mathbf{a}_y = 1}} \left[ (1-\alpha) \cdot \mathbf{a}_x^T \mathbf{Q}_x \mathbf{a}_x - \alpha \cdot \mathbf{a}_x^T \mathbf{B} \mathbf{a}_y \right].$$

## Dual function

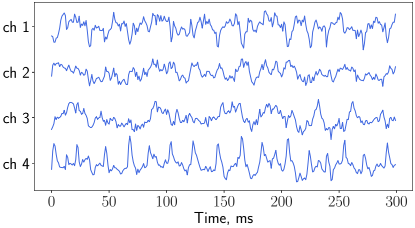
$$g(\mathbf{a}_x, \lambda, \boldsymbol{\mu}) = \begin{cases} (1-\alpha) \cdot \mathbf{a}_x^T \mathbf{Q}_x \mathbf{a}_x - \lambda, & \alpha \cdot \mathbf{B}^T \mathbf{a}_x = \lambda \cdot \mathbf{1}_r + \boldsymbol{\mu}; \\ +\infty, & \text{otherwise.} \end{cases}$$

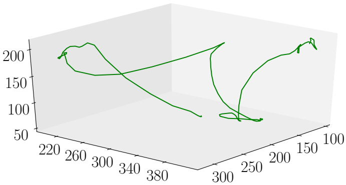
## Statement

*For the case $r = 1$ the proposed strategies coincide with the original QPFS algorithm.*

# Computational experiment

## Data example



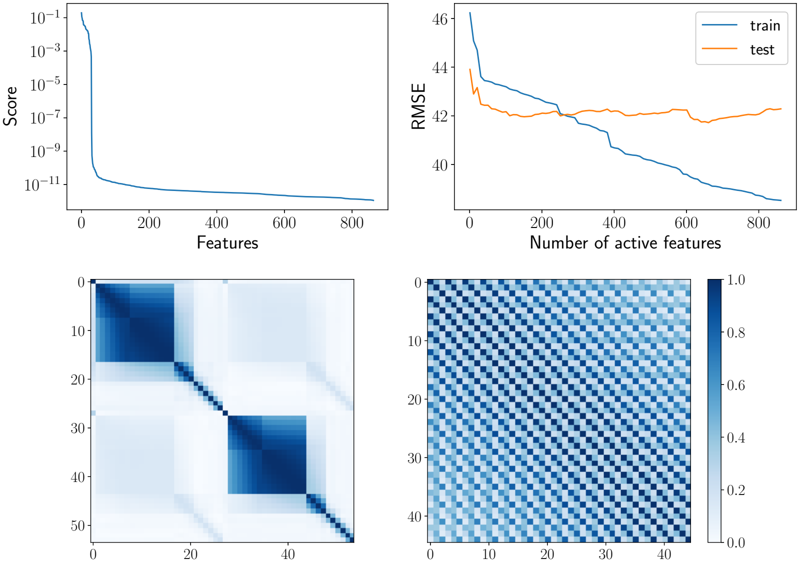




$$\mathbf{X} \in \mathbb{R}^{m\times(32\cdot 27)}; \quad \mathbf{Y} = \begin{pmatrix} x_1 & y_1 & z_1 & \dots & x_k & y_k & z_k \\ x_2 & y_2 & z_2 & \dots & x_{k+1} & y_{k+1} & z_{k+1} \\ & \dots & & \dots & & \dots & \\ x_m & y_m & z_m & \dots & x_{m+k} & y_{m+k} & z_{m+k} \end{pmatrix} \in \mathbb{R}^{m\times 3k}.$$

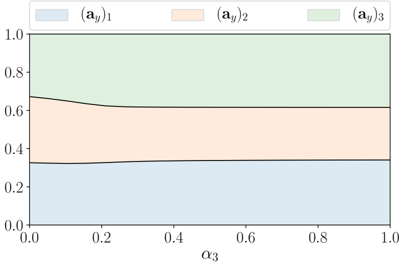
---

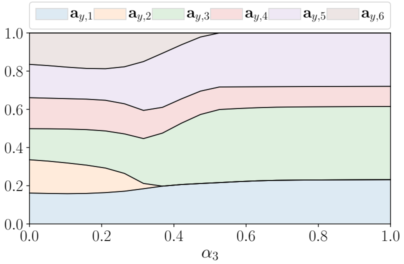
http://neurotycho.org

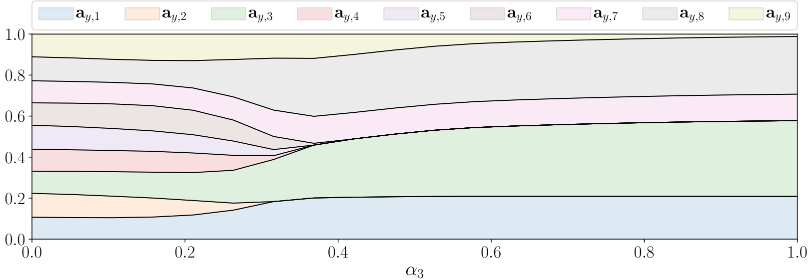
# Data redundancy



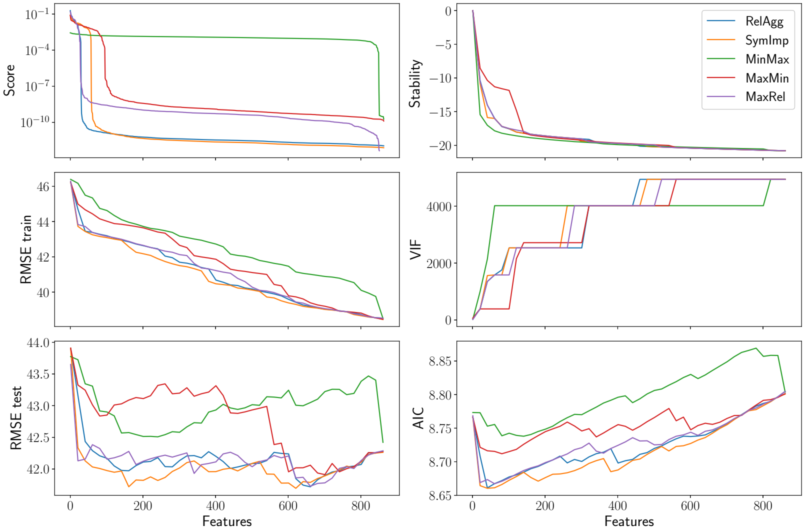
# Target importances



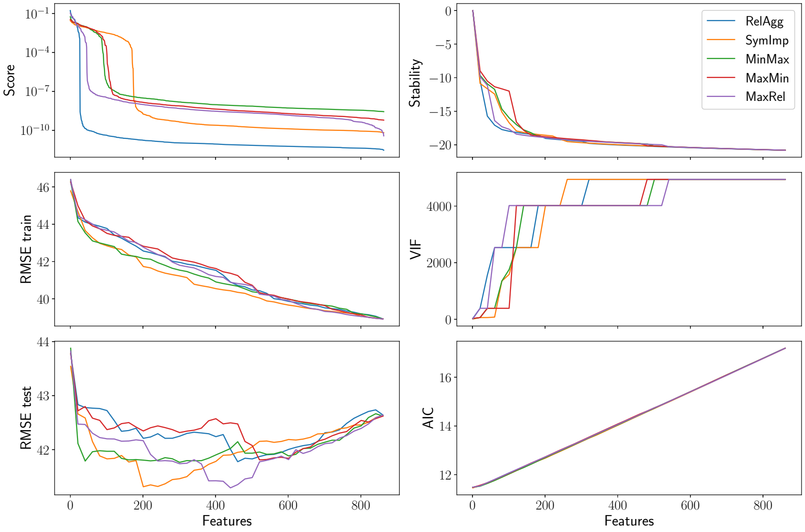



# Autoregression step $= 1$



# Autoregression step = 30



# Conclusion

- BCI signal decoding problem is investigated.
- Feature selection algorithms for spatio-temporal data are proposed.
- Suggested algorithms are explored and compared.

## Publications

- Isachenko R. et al. Feature Generation for Physical Activity Classification. *Artificial Intellegence and decision making*, submitted.
- Isachenko R., Strijov V. Quadratic programming optimization for Newton method. *Lobachevskii Journal of Mathematics*, submitted.
- Isachenko R., Vladimirova M., Strijov V. Dimensionality reduction for time series decoding and forecasting problems. Ready for submission.

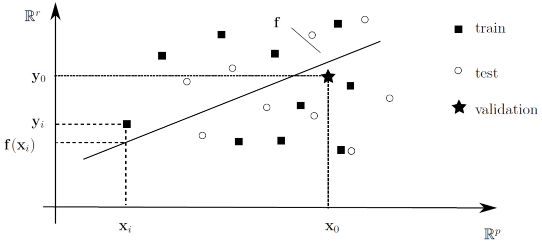
## Feature categorization

1. non-relevant features
$$\left\{j : \mathrm{corr}(\boldsymbol{\chi}_j, \boldsymbol{\nu}_k) = 0, \ \forall k \in \{1, \ldots, r\}\right\};$$
2. non-**X**-correlated features, which are relevant to non-**Y**-correlated targets
$$\left\{j : \left(\mathrm{VIF}(\boldsymbol{\chi}_j) < 10\right) \text{ and } \left(\mathrm{VIF}(\boldsymbol{\nu}_k) < 10, \ \forall k \in \{1, \ldots, r\} : \mathrm{corr}(\boldsymbol{\chi}_j, \boldsymbol{\nu}_k) \neq 0\right)\right\};$$
3. non-**X**-correlated features, which are relevant to **Y**-correlated targets
$$\left\{j : \left(\mathrm{VIF}(\boldsymbol{\chi}_j) < 10\right) \text{ and } \left(\exists k \in \{1, \ldots, r\} : \mathrm{VIF}(\boldsymbol{\nu}_k) > 10 \ \& \ \mathrm{corr}(\boldsymbol{\chi}_j, \boldsymbol{\nu}_k) \neq 0\right)\right\};$$
4. **X**-correlated features, which are relevant to non-**Y**-correlated targets
$$\left\{j : \left(\mathrm{VIF}(\boldsymbol{\chi}_j) > 10\right) \text{ and } \left(\mathrm{VIF}(\boldsymbol{\nu}_k) < 10, \ \forall k \in \{1, \ldots, r\} : \mathrm{corr}(\boldsymbol{\chi}_j, \boldsymbol{\nu}_k) \neq 0\right)\right\};$$
5. **X**-correlated features, which are relevant to **Y**-correlated targets
$$\left\{j : \left(\mathrm{VIF}(\boldsymbol{\chi}_j) > 10\right) \text{ and } \left(\exists k \in \{1, \ldots, r\} : \mathrm{VIF}(\boldsymbol{\nu}_k) > 10 \ \& \ \mathrm{corr}(\boldsymbol{\chi}_j, \boldsymbol{\nu}_k) \neq 0\right)\right\}.$$

$$r = \mathrm{corr}(\boldsymbol{\chi}, \boldsymbol{\nu}), \quad t = \frac{r\sqrt{m-2}}{1-r^2} \sim \mathrm{St}(m-2).$$
$$\mathrm{VIF}(\boldsymbol{\chi}_j) = \frac{1}{1-R_j^2}, \quad \mathrm{VIF}(\boldsymbol{\nu}_k) = \frac{1}{1-R_k^2},$$

where $R_j^2(R_k^2)$ are coefficients of determination for the regression of $\boldsymbol{\chi}_j(\boldsymbol{\nu}_k)$ on the other features(targets).

## Problem Statement




$$
\begin{tikzcd}
\mathbf{X} \in \mathbb{X} \arrow[r] \arrow[d, leftrightarrow] & \mathbf{Y} \in \mathbb{Y} \arrow[d, leftrightarrow] \\
\mathbf{T} \in \mathbb{Z} \arrow[r, leftrightarrow] & \mathbf{U} \in \mathbb{Z}
\end{tikzcd}
$$

### Partial Least Squares (PLS)

$$\underset{m\times n}{\mathbf{X}} = \underset{m\times l}{\mathbf{T}} \cdot \underset{l\times n}{\mathbf{P}^T} + \underset{m\times n}{\mathbf{F}} = \sum_{k=1}^{l} \underset{m\times 1}{\mathbf{t}_k} \cdot \underset{1\times n}{\mathbf{p}_k^T} + \underset{m\times n}{\mathbf{F}}$$

$$\underset{m\times r}{\mathbf{Y}} = \underset{m\times l}{\mathbf{U}} \cdot \underset{l\times r}{\mathbf{Q}^T} + \underset{m\times r}{\mathbf{E}} = \sum_{k=1}^{l} \underset{m\times 1}{\mathbf{t}_k} \cdot \underset{1\times r}{\mathbf{q}_k^T} + \underset{m\times r}{\mathbf{E}}$$

$$\hat{\mathbf{Y}} = \mathbf{T}\mathrm{diag}(\beta)\mathbf{Q}^T = \mathbf{X}\boldsymbol{\Theta}.$$

- map $\mathbf{X}$ into low-dimensional $\mathbf{T}$;
- map $\mathbf{Y}$ into low-dimensional $\mathbf{U}$;
- maximize correlation between $\mathbf{t}_k$ and $\mathbf{u}_k$.

# PLS Example

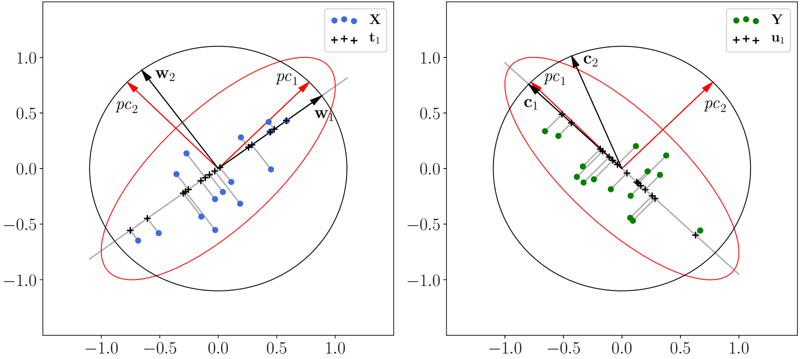



Figure: The result of the PLS algorithm for the case $n = r = l = 2$.

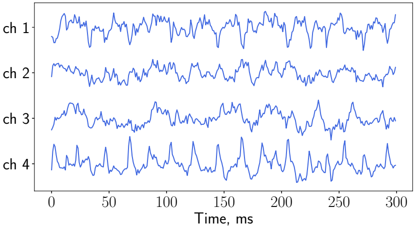
## Datasets

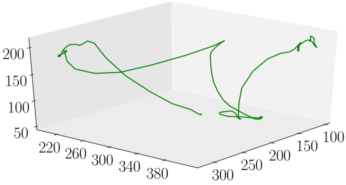
- energy consumption
- electrocorticogram signals (ECoG)

## Autoregressive approach

$$\mathbf{X} = \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ x_2 & x_3 & \dots & x_{n+1} \\ \dots & \dots & \dots & \dots \\ x_{T-n+1} & x_{T-n+2} & \dots & x_T \end{pmatrix}$$

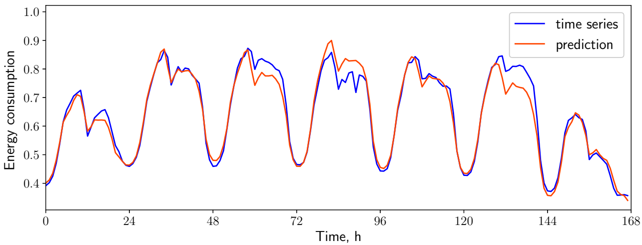
## ECoG data



# Computational experiment

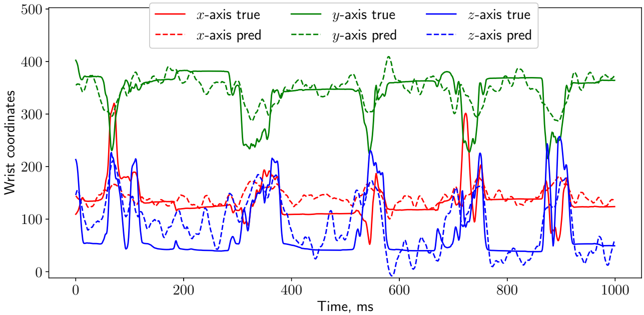
## Energy consumption




## Results

- Space dimensionalities: $\mathbf{X} = 700 \times (24 \cdot 7)$ , $\mathbf{Y} = 700 \times 24$.
- Dimensionality of latent space: 14
- NMSE: 0.047

## ECoG




## Results

- Space dimensionalities: $\mathbf{X} = 13000 \times (864 \cdot 18)$ , $\mathbf{Y} = 13000 \times 3$.
- Dimensionality of latent space: 16
- NMSE: 0.731

# Quadratic Programming Model Selection

## QPFS

- works for linear problems;
- does not take into account the model;
- ignores the structure of the target space.

## Problem

$$\underbrace{(1-\alpha)\mathbf{z}^{T}\mathbf{Q}\mathbf{z}}_{\text{Sim}} - \underbrace{\alpha\mathbf{b}^{T}\mathbf{z}}_{\text{Rel}} \to \min_{\substack{\mathbf{z}\in\mathbb{R}^{p}_{+} \\ \|\mathbf{z}\|_1=1}} .$$

- $\mathbf{z} \in \mathbb{R}^p$ — weight importances;
- $\mathbf{Q} \in \mathbb{R}^{p\times p}$ - pairwise weights interactions;
- $\mathbf{b} \in \mathbb{R}^p$ - weight relevances to the target vector.

$$w_j = 0 \Leftrightarrow z_j < \tau.$$