Supplementary material to ‘Statistical Ranking of Electromechanical Dyssynchrony parameters for CRT’

Matthieu Toulemon and Julien Reygner

ABSTRACT. This document presents the mathematical and algorithmic details of the variance-based variable selection procedure employed in the statistical analysis of the Electromechanical Dyssynchrony parameters for Cardiac Resynchronization Therapy presented in [1].

1. Introduction

During the Meteor Study, whose results are presented and analyzed in [1], \( n = 455 \) experiments were conducted, each of which provided a measurement of the values of \( p = 18 \) different dyssynchrony parameters. The empirical means and standard deviations of this sample are reported in Table 1. Its correlation matrix is reproduced in Table 2.

| Name of variable | Mean | Std  | Unit |
|------------------|------|------|------|
| DFT%             | 0.44 | 0.078| None |
| QRS-E            | 521.79 | 60.714 | s    |
| RPEI             | 128.77 | 38.648 | s    |
| LPEI             | 161.98 | 39.33 | s    |
| DIV              | 34.43 | 36.295 | s    |
| SD               | 447.6 | 47.238 | s    |
| Sept             | 553.33 | 116.231 | s    |
| LLW              | 495.23 | 88.755 | s    |
| DiastC sept      | 106.64 | 103.881 | s    |
| DiastClat        | 48.22 | 84.519 | s    |
| IsovolRT         | 71.74 | 51.81 | s    |
| OverlapSept      | 36.11 | 117.692 | s    |
| OverlapLat       | -22.35 | 89.006 | s    |
| Sept-LLW         | 58.04 | 121.176 | s    |
| MVRsurf/LA       | 0.22 | 0.184 | None |
| IsovolCT         | 102.95 | 75.422 | s    |
| LVET             | 285.99 | 45.181 | s    |
| LPEI/LVET        | 0.59 | 0.2 | s    |

TABLE 1. Empirical means and standard deviations for the Meteor Study.
Table 2. Empirical correlations for the Meteor Study. Blank cells correspond to pairs of variables for which the Pearson correlation test indicates that the variables are not correlated, with a $p$-value lower than 0.001.
Measuring the value of each of these parameters brings forth an important experimental cost; for instance, implanting a sensor in the patient’s heart. On the other hand, since the Meteor Study evidenced that several parameters are strongly correlated to each other, the information brought by the supplementary measurement of a parameter which is highly correlated with an already measured parameter may be limited. Therefore, one may imagine to reduce the cost of the experiment by measuring only a few parameters, chosen in order to minimize the loss of information. The purpose of this note is to quantify this idea.

In the sequel, we let \( \mathbf{X} = (X_1, \ldots, X_p)^\top \) be a random vector, with expectation \( \mathbf{\mu} = (\mu_1, \ldots, \mu_p)^\top \) and covariance matrix \( \mathbf{K} \), whose coefficients are denoted by \( K_{i,j} \), \( 1 \leq i, j \leq p \). We shall work under the following assumptions.

(i) \( \mathbf{X} \) is a Gaussian vector.

(ii) The covariance matrix \( \mathbf{K} \) is invertible.

Using the notion of conditional distribution, we shall look for a partition of the set of indices \( \{1, \ldots, p\} \) into two nonempty sets \( \mathcal{O} \) (for the set of observed variables) and \( \mathcal{U} \) (for the set of unobserved variables) such that the observation of the values of the variables \( X_i, i \in \mathcal{O} \) is the most informative on the values of the variables \( X_j, j \in \mathcal{U} \).

2. Description of the method

2.1. Conditional confidence intervals. Let us fix a partition of the set of indices \( \{1, \ldots, p\} \) into two nonempty sets \( \mathcal{O} \) and \( \mathcal{U} \). We introduce the block decompositions

\[
(1) \quad \mathbf{X} = \begin{pmatrix} \mathbf{X}_\mathcal{O} \\ \mathbf{X}_\mathcal{U} \end{pmatrix}, \quad \mathbf{\mu} = \begin{pmatrix} \mathbf{\mu}_\mathcal{O} \\ \mathbf{\mu}_\mathcal{U} \end{pmatrix}, \quad \mathbf{K} = \begin{pmatrix} \mathbf{K}_\mathcal{O} & \mathbf{K}_{\mathcal{O},\mathcal{U}} \\ \mathbf{K}_{\mathcal{U},\mathcal{O}} & \mathbf{K}_\mathcal{U} \end{pmatrix},
\]

of \( \mathbf{X}, \mathbf{\mu} \) and \( \mathbf{K} \). The assumption that \( \mathbf{K} \) is invertible implies that \( \mathbf{K}_\mathcal{O} \) and \( \mathbf{K}_\mathcal{U} \) are invertible.

It is known [2, Proposition 3.13, p. 116] that the conditional law of \( \mathbf{X}_\mathcal{U} \) given the value of \( \mathbf{X}_\mathcal{O} \) is the Gaussian measure on \( \mathbb{R}^\mathcal{U} \) with expectation

\[
(2) \quad \mathbf{\mu}_{\mathcal{U}|\mathcal{O}}(\mathbf{X}_\mathcal{O}) = \mathbf{\mu}_\mathcal{U} + \mathbf{K}_{\mathcal{U},\mathcal{O}} \mathbf{K}_\mathcal{O}^{-1} (\mathbf{X}_\mathcal{O} - \mathbf{\mu}_\mathcal{O}),
\]

and covariance matrix

\[
(3) \quad \mathbf{K}_{\mathcal{U}|\mathcal{O}} = \mathbf{K}_\mathcal{U} - \mathbf{K}_{\mathcal{U},\mathcal{O}} \mathbf{K}_\mathcal{O}^{-1} \mathbf{K}_{\mathcal{O},\mathcal{U}},
\]

which has the remarkable property not to depend on the value of \( \mathbf{X}_\mathcal{O} \). For \( i \in \mathcal{U} \), the \( i \)-th coordinate of \( \mathbf{\mu}_{\mathcal{U}|\mathcal{O}}(\mathbf{X}_\mathcal{O}) \) is denoted by \( \mu_{i|\mathcal{O}}(\mathbf{X}_\mathcal{O}) \); for \( i, j \in \mathcal{U} \), the \( i, j \)-th coordinate of \( \mathbf{K}_{\mathcal{U}|\mathcal{O}} \) is denoted by \( K_{i,j|\mathcal{O}} \).

Fix \( \alpha \in (0, 1/2) \) and denote by \( \phi_{1-\alpha/2} \) the quantile of order \( 1 - \alpha/2 \) of the standard Gaussian distribution on \( \mathbb{R} \). As a consequence of the previous discussion, for any \( i \in \mathcal{U} \) the interval

\[
(4) \quad \mathcal{I}_{i|\mathcal{O}}(\mathbf{X}_\mathcal{O}) = \left[ \mu_{i|\mathcal{O}}(\mathbf{X}_\mathcal{O}) - \phi_{1-\alpha/2} \sqrt{K_{i,i|\mathcal{O}}}, \mu_{i|\mathcal{O}}(\mathbf{X}_\mathcal{O}) + \phi_{1-\alpha/2} \sqrt{K_{i,i|\mathcal{O}}} \right]
\]

is such that \( \mathbb{P}(X_i \in \mathcal{I}_{i|\mathcal{O}}(\mathbf{X}_\mathcal{O})|\mathbf{X}_\mathcal{O}) = 1 - \alpha \). The width of this interval is \( 2\phi_{1-\alpha/2} \sqrt{K_{i,i|\mathcal{O}}} \), so that the smaller the ratio \( \sqrt{K_{i,i|\mathcal{O}}}/\sqrt{K_{i,i}} \), the more information on the value of \( X_i \) is brought by the observation of \( \mathbf{X}_\mathcal{O} \). Let us denote by

\[
(5) \quad V_0 = \sum_{i \in \mathcal{U}} \frac{K_{i,i|\mathcal{O}}}{K_{i,i}}
\]

the sum of squares of these ratios, and take this quantity as the measure of global uncertainty remaining after the observation of \( \mathbf{X}_\mathcal{O} \). Notice that, in the extremal cases \( \mathcal{O} = \emptyset \) and \( \alpha = \{1, \ldots, p\} \), \( V_0 \) takes the respective values \( p \) and 0.
2.2. Interpretation of $V_O$. The quantity $V_O$ can be rewritten in terms of the correlation matrix $C$ of $X$, the coefficients of which are defined by

$$\forall i, j \in \{1, \ldots, p\}, \quad C_{i,j} = \frac{K_{i,j}}{\sqrt{K_{i,i}K_{j,j}}}.$$  

Introducing the block decomposition

$$C = \begin{pmatrix} C_O & C_{U,O}^\top \\ C_{U,O} & C_U \end{pmatrix}$$  

and defining the conditional correlation matrix

$$C_{U|O} = C_U - C_{U,O}C_O^{-1}C_{U,O}^\top,$$

we claim that for all $i \in U$, we have

$$\frac{K_{i,i|O}}{K_{i,i}} = C_{i,i|O},$$

so that

$$V_O = \sum_{i \in U} C_{i,i|O} = \text{tr} \ C_{U|O}.$$  

This discussion leads us to adopt the following convention.

**Definition 2.1** (Global uncertainty remaining after an observation). Let $O$ be a nonempty strict subset of $\{1, \ldots, p\}$, and let $U = \{1, \ldots, p\} \setminus O$. The global uncertainty remaining after the observation of $X_O$ is the quantity $V_O = \text{tr} \ C_{U|O}.$

Turning back to the example of the Meteor Study, a possible approach to the optimal reduction of the experimental cost can be summarized as follows.

(i) Fix the number $q \in \{1, \ldots, p-1\}$ of variables that you want to observe, depending on the targeted experimental cost.

(ii) Select the subset $O \subset \{1, \ldots, p\}$ with cardinality $q$ which minimizes $V_O$.

(iii) Observe the value of $X_O$ and return the confidence intervals $I_{i|O}(X_O)$, $i \in U$, defined by (4).

2.3. Empirical covariance matrix. In practice, $\mu$, $K$, and $C$ are not known but rather estimated from a preliminary data collection. Using the properties of Wishart matrices [2, Chapter 8], it is theoretically possible to adapt the formula (4) to take the associated sampling error into account in the computation of the confidence interval. We shall not address this problem here and rather focus on the algorithmic determination of the optimal subset $O$ in Step (ii) of the approach described above. To this aim, we remark that working with empirical quantities entails a convenient reformulation.

Let $x_1, \ldots, x_n \in \mathbb{R}^p$ be the observed dataset, and denote by $A \in \mathbb{R}^{n \times p}$ the matrix containing the values of this data. We assume that these data have been standardized, so that the empirical mean and variance of each column of the matrix $A$ respectively equal 0 and 1. In this context, the correlation matrix $C$ is estimated by the matrix

$$\hat{C} = A^\top A,$$

from which we define $\hat{C}_O$, $\hat{C}_{U,O}$, $\hat{C}_U$ and $\hat{C}_{U|O}$ in a straightforward fashion. On account of Definition 2.1, the selection of the subset $O$ reduces to the minimization of $\text{tr} \ \hat{C}_{U|O}$. We now show that this problem is equivalent to the so-called Column Subset Selection problem (also $CU$ factorization) in Linear Algebra and Machine Learning [3, 4].
Lemma 2.2 (Equivalence with Column Subset Selection problem). Let \( \{O, U\} \) be a partition of \( \{1, \ldots, q\} \) into two nonempty subsets. Denote by

\[
A = (A_O \quad A_U)
\]

the corresponding block decomposition of \( A \), and let \( \Pi_O \) denote the orthogonal projection of \( \mathbb{R}^n \) onto the subspace spanned by the columns of \( A_O \). Then

\[
\text{tr} \hat{C}_{\mid O} = \| A - \Pi_O A \|^2,
\]

where \( \| \cdot \| \) denotes the Frobenius norm on \( \mathbb{R}^{n \times p} \).

Proof. By the definition of the Frobenius norm,

\[
\| A - \Pi_O A \|^2 = \text{tr} \left((A - \Pi_O A)^T (A - \Pi_O A)\right) = \text{tr} \left(A^T A - A^T \Pi_O A\right).
\]

The orthogonal projection of \( \mathbb{R}^n \) onto the subspace spanned by the columns of \( A_O \) writes

\[
\Pi_O = A_O A_O^+
\]

where

\[
A_O^+ = (A_O^T A_O)^{-1} A_O^T
\]

is the Moore-Penrose inverse of \( A_O \). We deduce that

\[
\| A - \Pi_O A \|^2 = \text{tr} \left(A^T A - A^T A_O (A_O^T A_O)^{-1} A_O^T A\right).
\]

Using the block decomposition of \( A \), we get

\[
A^T A_O (A_O^T A_O)^{-1} A_O^T A = \begin{pmatrix}
A_O^T A_O & A_O^T A_U \\
A_U^T A_O & A_U^T (A_O^T A_O)^{-1} A_U
\end{pmatrix},
\]

so that

\[
\text{tr} \left(A^T A - A^T A_O (A_O^T A_O)^{-1} A_O^T A\right) = \text{tr} \left(A_U^T A_U - A_U^T A_O (A_O^T A_O)^{-1} A_O^T A_U\right).
\]

On the other hand, the block decomposition of \( A \) also provides the identity

\[
\hat{C} = \begin{pmatrix}
\hat{C}_O & \hat{C}_{U, O}^T \\
\hat{C}_{U, O} & \hat{C}_U
\end{pmatrix} = \begin{pmatrix}
A_O^T A_O & A_O^T A_U \\
A_U^T A_O & A_U^T A_U
\end{pmatrix},
\]

from which we deduce that

\[
\hat{C}_{\mid O} = \hat{C}_U - \hat{C}_{U, O} \hat{C}_O^{-1} \hat{C}_{U, O}^T = A_U^T A_U - A_U^T A_O (A_O^T A_O)^{-1} A_O^T A_U,
\]

which completes the proof. \(\square\)

The exact resolution of the problem would imply computing \( \hat{K}_{\mid O} \) for each possible subset of the 18 variables and for each subset size. This resolution is not tractable when the number of variables is too large, in fact [4] shows that the Column Subset Selection problem is NP-complete. In order to solve it approximately in a reasonable amount of time we use a greedy method described in the next section.
3. Numerical experiments for the Meteor Study

For the Meteor Study, our dataset has $n = 455$ rows and $p = 18$ columns. We recall that the empirical mean and standard deviation of the $p$ parameters are reported in Table 1. Denoting by $A \in \mathbb{R}^{n \times p}$ the standardized dataset, our purpose is now to select, for $q \in \{1, \ldots, p-1\}$, the subset $O \subset \{1, \ldots, p\}$ with cardinality $q$ which minimizes $\|A - \Pi_O A\|^2$.

To do so, we apply the greedy algorithm described in [3], which works as follows.

Algorithm 1 Greedy Column Subset Selection

Require: initial index $i_0 \in \{1, \ldots, p\}$
\begin{align*}
O &\leftarrow \{i_0\} \\
U &\leftarrow \{1, \ldots, p\} \setminus \{i_0\}
\end{align*}
while $U \neq \emptyset$ do
\begin{align*}
&\text{find } j \in U \text{ which minimizes } \|A - \Pi_{O \cup \{j\}} A\|^2 \\
&\text{add } j \text{ to } O \\
&\text{remove } j \text{ from } U
\end{align*}
end while

As the number of variables in the Meteor Study was relatively small we considered all possible initializations $i_0 \in \{1, \ldots, p\}$. We thus obtained $p$ sequences of increasing sets
\[
\{i_0\} = O^{(i_0)}_1 \subseteq O^{(i_0)}_2 \subseteq \cdots \subseteq O^{(i_0)}_p = \{1, \ldots, p\}, \quad i_0 = 1, \ldots, p,
\]
from which we selected, for any $q \in \{1, \ldots, p-1\}$, the subset
\[
O^*_q \in \{O^{(i_0)}_q, i_0 = 1, \ldots, p\}
\]
which minimized $\|A - \Pi_O A\|^2$.

The sequence of obtained subsets $\{O^*_q, q = 1, \ldots, p\}$ is represented on Table 3, together with the associated proportion of reduced variance $1 - V_{O^*_q}/p$. For visualization purposes $V_{O^*_q}/p$ was rounded up to 0.001, therefore the last two lines should be interpreted as $V_{O^*_q}/p \leq 0.001$.

Remark 3.1. We observed that for small values of $q$, the choice of the initial index $i_0$ is important. This is due to the large number of variables strongly correlated to Sept in the dataset (OverlapSept, DiastCSept, . . .). For subsets containing more than 5 variables, it seems that the choice of the initial index becomes less important as many groups achieve about the same reduction of uncertainty.

Acknowledgements. We wish to thank Guillaume Obozinski for fruitful discussions.

References

[1] S. Cazeau, M. Toulemont, P. Ritter, and J. Reygner (2018). Statistical Ranking of Electromechanical Dyssynchrony parameters for CRT. Submitted.
[2] M. L. Eaton (2007). Multivariate Statistics: A Vector Space Approach. Institute of Mathematical Statistics.
[3] A.K. Farahat, A. Ghodsi, and M. S. Kamel (2013). A Fast Greedy Algorithm for Generalized Column Subset Selection. NIPS’13 Workshop on Greedy Algorithms, Frank-Wolfe and Friends.
[4] Y. Shitov (2017). Column Subset Selection is NP-complete.

Matthieu Toulemont
Département Ingénierie Mathématique et Informatique, École des Ponts ParisTech, Marne-la-Vallée, France

Julien Reygner
Université Paris-Est, CERMICS (ENPC), Marne-la-Vallée, France
| q | Sept | 0.239 | 0.416 | 0.564 | 0.707 | 0.786 | 0.859 | 0.919 | 0.958 | 0.984 | 0.990 | 0.996 | 0.998 | 0.999 | 1.000 |
|---|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 1 | Sept |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 2 | QRS-E | Sept |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 3 | QRS-E |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 4 | QRS-E |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 5 | DFT% | Sept |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 6 | LPEI | Sept |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 7 | LVET | Sept |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 8 | LVET |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 9 | LVET |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 10 | SD |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 11 | SD |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 12 | SD |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 13 | SD |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 14 | SD |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 15 | SD |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 16 | SD |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 17 | SD |      |      |      |      |      |      |      |      |      |      |      |      |      |      |

Each line represents the best group for a given number of variables to select, at the end of each line is written 1-V.