Report on “A mathematical programming approach to sparse canonical correlation analysis”

Submitted to European Journal of Operational Research

January 13, 2023

In this paper, the authors considered the sparse canonical correlation analysis (CCA) problem:

$$\begin{aligned}\max_{w_1,w_2} \quad & w_1^T X_1^T X_2 w_2 \\ \text{s.t.} \quad & w_1^T X_1^T X_1 w_1 = w_2^T X_2^T X_2 w_2 = 1, \\ & \|w_1\|_0 \le k_1, \quad \|w_2\|_0 \le k_2,\end{aligned}$$

where X_1 and X_2 are data matrices, k_1 and k_2 are integers and $\|\cdot\|_0$ is the cardinality function. The authors reformulated the above problem into the following mixed integer quadratic programming problem:

$$\begin{aligned}\max_{w_1,w_2,z_1,z_2} \quad & w_1^T X_1^T X_2 w_2 \\ \text{s.t.} \quad & w_1^T X_1^T X_1 w_1 = w_2^T X_2^T X_2 w_2 = 1, \\ & -C_j z_{j,i} \le w_{j,i} \le C_j z_{j,i}, \quad i = 1,\ldots,p_j \quad j = 1,2, \\ & z_{j,i} \in \{0,1\} \quad i = 1,\ldots,p_j \quad j = 1,2, \\ & \textstyle\sum_{i=1}^{p_j} z_{j,i} \le k_j \quad j = 1,2,\end{aligned}$$

where C_1 and C_2 are sufficiently large positive constants. They then adapted existing methods such as alternating minimization (Section 3.2), a bender-like decomposition method (Section 3.3), a kernel search method (Section 3.4) and a hybrid method that combines the alternating minimization method and kernel search method (Section 3.5) to solve the above model. Numerical experiments are performed comparing the proposed approaches against existing methods such as the “PMA” R-package.

From what I see, the advantage of the best approach proposed in this paper (the method in Section 3.5) is not very obvious. The authors also did not emphasize enough the novelty in their adaptation of existing MINLP methods. Furthermore, the descriptions of some of the approaches, especially the validity of the bender-like decomposition method, also need some clarifications. I think the overall contribution may not merit publication in EJOR. I would suggest a **rejection**.

Major comments:

1. I am confused by the validity of the derivation of the bender-like decomposition in Section 3.3. The problem (25)-(30) looks like the Wolfe dual problem to me, and in general there will be duality gap because the problem is nonconvex. I am thus not sure about the argument around the last display of the proof of Theorem 3.2, where the authors claim that $W = \min_{\alpha,\beta} \max_{w,z} \mathcal{L}$, and that the equality holds in the display. If there are standard references, please provide a citation. Otherwise, please argue carefully about the validity of the procedure.

2. I suppose that in Figure 1, the higher the dots, the better the results. Please say this explicitly in the text. As for the comparison between BRKS and l1-rg1, I can see from Figure 2 that l1-rg1 is always faster. So it does not really surprise me that the dots obtained by l1-rg1 in Figure 1 are worse. For a fair comparison, I think the authors may also plot the heights of the dots as the algorithms BRKS and l1-rg1 progress, to illustrate the performances of these algorithms over time.

Other comments:

1. In (7), please write "$j = 1, 2$" explicitly. Moreover, this problem is not convex because of (6). Thus, please rephrase the sentence right below (7).

2. Right after the first display of the proof of Theorem 3.2: $\gamma \in \mathbb{R}$ instead of being nonnegative. And, since (26) consists of equality constraints, the third relation in the third display of the proof of Theorem 3.2 should be removed.

3. In Algorithm 2, line 1, what precisely are the authors considering as the "continuous relaxation problem"? And I also wonder why the procedure described in lines 2–6 in Algorithm 2 is not used in Algorithm 3.

4. In Section 4, line 10, by "exact solution of the SCCA formulation", do the authors mean solving the MIQO problem in Section 3.1? And please explain what is meant by BR_Ms.

Typos:

1. Please make a new paragraph starting from "The remainder..." in line -6 of the introduction.

2. Line 17 of the second paragraph of Section 2: "sparsity" instead of "sparseness".

3. Line 2 of Section 3: please delete "where".

4. Line 8 of Section 3: please add "the" before "transformed variables".

5. Last line on page 6: "reduce to transforming".

6. The line below (22): "it is impossible".