 Subset selection in sparse matrices
Alberto Del Pia † Santanu S. Dey ‡ Robert Weismantel †
October 5, 2018

Abstract
In subset selection we search for the best linear predictor that involves a small subset of variables. From a computational complexity viewpoint, subset selection is NP-hard and few classes are known to be solvable in polynomial time. Using mainly tools from discrete geometry, we show that some sparsity conditions on the original data matrix allow us to solve the problem in polynomial time.

Key words: subset selection, linear regression, polynomial-time algorithm, sparsity

1 Introduction
In machine learning and statistics, subset selection is also known as feature selection, attribute selection, variable selection or variable subset selection. It is the problem of selecting a subset of relevant variables (or features) to recover a predictor variable. Subset selection techniques are used for three main reasons: (i) Improve prediction accuracy by reducing the variance of the predicted values (reducing overfitting); (ii) Simplify the model to make it easier to interpret; (iii) Decrease prediction times, since only few variables must be sampled every time a prediction is required. In subset selection, the prior knowledge is that the data contains many variables that are either redundant or irrelevant, and can thus be removed without incurring much loss of information. Natural applications of subset selection abound in medical or social studies. As an example, consider the problem of predicting the risks of heart disease in terms of observable quantities such as age, sex, blood pressure, cholesterol level, etc. The goal is to identify a small set of attributes for future tests.

Due to the vast applicability of this model, many approaches have been proposed by different communities, including greedy algorithms (e.g., forward- and backward-stepwise selection [16, 8, 4], forward-stagewise regression [7, 11]), branch and bound [2, 13] (e.g., the leaps and bounds procedure [9]), and convex optimization (e.g., ridge regression [14], the lasso [17]). See [11] for an introduction to subset selection.

Formally, subset selection is a nonlinear optimization problem of the following form:

\[
\begin{align*}
\min & \quad \| Mx + c\mu - b \| \\
\text{s.t.} & \quad x \in \mathbb{R}^d, \quad \mu \in \mathbb{R} \\
& \quad |\text{supp}(x)| \leq \sigma.
\end{align*}
\]  

\(1\)
In this formulation, $x$ is the $d$-vector of unknowns and $\mu$ is a scalar variable. The remaining characters stand for data in the problem instance: $M$ is an $m \times d$ matrix, $b$ and $c$ are $m$-vectors, and $\sigma$ is a natural number. Finally, $\|\cdot\|$ denotes the Euclidean norm. Note that in standard formulations of the subset selection problem it is often assumed that $c$ is the vector of all ones. It is often assumed that the columns of $M$ and $b$ are mean-centered (i.e., the sum of entries in the columns of $M$ and in $b$ is zero), in which case it can be shown that the optimal value of $\mu$ is zero. Since in this paper we want to exploit the sparsity structure of $M$, we do not assume mean-centering and therefore explicitly retain the $\mu$ variable.

From a computational complexity point of view, subset selection is NP-hard [18]. Only few results on polynomially solvable cases regarding subset selection are known. Das and Kempe [1] give an exact algorithm when the covariance graph is a tree. The covariance graph has its nodes associated with the variables and edges between any pair of variables with non-zero covariance. Another special case has been analyzed by the same authors. If the covariance graph has a stable set of size number of variables minus a constant and if this stable set is explicitly known, then subset selection becomes polynomial time solvable. All these results have been obtained by exploiting matrix perturbation techniques. Das and Kempe [1] also present a fully polynomial time approximation scheme when the covariance matrix has constant bandwidth. Donoho [5] and Candes, Romberg and Tao [3] show that under mild conditions, replacing the cardinality constraint with an $l_1$-constraint yields the exact solution with an overwhelming probability. Gao and Li [10] give an exact algorithm when the $d-k$ largest eigenvalues of the matrix $M^T M$ are identical. Here, $M$ denotes the data matrix in Eq. (1). Moreover, this result requires that $k$ is a fixed number.

In this paper we are interested in identifying sparsity conditions on the original data matrix $M$ that allow us to solve subset selection in polynomial time. Our approach relies in contrast to all previously known polynomial time results on tools from discrete geometry and an analysis of the proximity of optimal solutions with respect to two consecutive “support-conditions” $|\text{supp}(x)| \leq s$ and $|\text{supp}(x)| \leq s+1$. Our main result is that subset selection can be solved in polynomial time if the matrix $M$ is obtained by adding a fixed number of extra columns to a block diagonal matrix, where each block involves a fixed number of variables. Therefore the matrices that we study have the following sparsity structure:

$$M = \begin{pmatrix} A^1 & \cdots & c_1 & \cdots & c_k \\ \vdots & \ddots & \vdots & \cdots & \vdots \\ A^h & \cdots & c_1 & \cdots & c_k \end{pmatrix}.$$ \hspace{1cm} (2)

This setting naturally occurs in a number of real-world applications. In the heart disease example described above, the columns of $M$ correspond to the observable quantities (or features), and each row of $M$ corresponds to a different patient. We can then partition all patients based on their nationality: the patients that have nationality $i$ are the ones corresponding to the rows of $A_i$. The patients of nationality $i$ have been all tested for a number of features: the ones tested only in their own country are the ones corresponding to the columns of $A_i$, and the ones that are common to all countries are the ones corresponding to the extra columns.

This paper is organized as follows. In Section 2 we introduce a reduction of problem (1) to separable form. In Section 3 we study the case where $M$ is obtained from a diagonal matrix by adding a fixed number of extra columns. In Section 4 we study the case where $M$ is obtained by adding a fixed number of extra columns to a block diagonal matrix, where each block involves a fixed number of variables. The polynomial-time algorithm presented in Section 3 clearly implies the polynomial solvability of the class discussed in Section 3. We present our algorithms in this order, since the proof for the diagonal case features many key ideas that will also be used in the block diagonal case, but with significantly more technicalities.
2 Reduction to separable form

All the data matrices \( M \) that we consider in this paper are of the form (2). Let \( n := d - k \), and denote by \( A \) the matrix

\[
A := \begin{pmatrix}
A^1 \\
\vdots \\
A^h
\end{pmatrix} \in \mathbb{R}^{m \times n},
\]

our matrix \( M \) can be written as \( M = (A|c_1| \cdots |c_k) \). With this notation, problem (1) takes the form

\[
\begin{align*}
\min & \quad \| (A|c_1| \cdots |c_k)x + c\mu - b \|
\text{s.t.} & \quad x \in \mathbb{R}^d, \mu \in \mathbb{R} \\
& \quad |\text{supp}(x)| \leq \sigma,
\end{align*}
\]

where \( b, c, c_1, \ldots, c_k \in \mathbb{R}^m \), and \( \sigma \in \mathbb{N} \).

In this section we introduce a reduction that will be useful throughout the paper. The reduced problem will have two key benefits over problem (1): (i) The support constraint will be applied only to the variables associated with the columns of \( A \); (ii) The objective function, once squared, will be decomposable in the subvectors of \( x \) corresponding to the blocks of \( A \). This however comes at the price of considering a fixed number of problems, instead of just one. Moreover, in each new problem, the one-dimensional parametric right hand side \( c\mu - b \) is replaced by a higher dimensional parametric right hand side. For the ease of notation, in the model (4) that we introduce in Lemma 1 below, the parameter \( k \) does not match the \( k \) in model (3); Namely, \( k \) in (4) is at most \( k + 1 \) in (3).

Lemma 1. Problem (3) can be polynomially reduced to a fixed number of problems of the form

\[
\begin{align*}
\min & \quad \left\| Ax - \left( b - \sum_{\ell=1}^{k} c_\ell \lambda_\ell \right) \right\| \\
\text{s.t.} & \quad x \in \mathbb{R}^n, \lambda \in \mathbb{R}^k \\
& \quad |\text{supp}(x)| \leq \sigma,
\end{align*}
\]

where \( A \in \mathbb{R}^{m \times n} \), \( b, c_1, \ldots, c_k \in \mathbb{R}^m \), \( \sigma \in \mathbb{N} \), and \( k \) is a fixed number.

Proof. Note that the objective function of problem (3) can be written in the form

\[
\left\| Ax - \left( b - c\mu - \sum_{\ell=1}^{k} c_\ell x_{n+\ell} \right) \right\|.
\]

For each subset \( \mathcal{L} \) of \( \{1, \ldots, k\} \), we consider the subproblem obtained from (3) by setting \( x_{n+\ell} = 0 \) for all \( \ell \in \{1, \ldots, k\} \setminus \mathcal{L} \), and by restricting the cardinality constraint to the \( n \)-dimensional vector \((x_1, \ldots, x_n)\). Formally,

\[
\begin{align*}
\min & \quad \left\| Ax - \left( b - c\mu - \sum_{\ell \in \mathcal{L}} c_\ell x_{n+\ell} \right) \right\| \\
\text{s.t.} & \quad x_1, \ldots, x_n \in \mathbb{R}, x_{n+\ell} \in \mathbb{R}, \forall \ell \in \mathcal{L}, \mu \in \mathbb{R} \\
& \quad |\text{supp}(x_1, \ldots, x_n)| \leq \sigma - |\mathcal{L}|.
\end{align*}
\]
To solve problem (3) we just need to solve the $2^k$ distinct subproblems of the form (5). To see this, let $(x^*, \mu^*)$ be an optimal solution of (3). Consider the subproblem (5) corresponding to the set $L := \{\ell \in \{1, \ldots, k\} : x^*_n + \ell \neq 0\}$, and let $(\hat{x}, \hat{\mu})$ be an optimal solution. Since the restriction of $(x^*, \mu^*)$ obtained by dropping the zero components $x^*_n + \ell$, for $\ell \in \{1, \ldots, k\} \setminus L$, is feasible for this subproblem, we have that the objective value of $(\hat{x}, \hat{\mu})$ is at most that of $(x^*, \mu^*)$. Consider now the extension of $(\hat{x}, \hat{\mu})$ obtained by adding components $\hat{x}_{n+\ell} = 0$ for all $\ell \in \{1, \ldots, k\} \setminus L$. This vector is feasible for (3), thus it is an optimal solution to the original problem (3). Since the extension of a feasible solution of each subproblem (5) is feasible for (3), we have that the best of the $2^k$ optimal solutions of the subproblems (5) will be an optimal solution to the original problem (3).

Since $k$ is a fixed number, also $|L|$ is a fixed number for any choice of $L$. By redefining $k := |L| + 1$, by introducing a $k$-dimensional vector $\lambda$ of variables $\mu$ and $x_{n+\ell}$, for $\ell \in L$, by redefining accordingly the vectors $c_\ell$, and redefining $\sigma := \sigma - |L|$, each subproblem (5) can be written in the form (3).

3 The diagonal case

In this section we consider problem (1), where the matrix $M$ is obtained from a diagonal matrix by adding $k$ extra columns. Hence, we consider matrices $M$ of the form (2) where each $A_i$ is a $1 \times 1$-matrix. In this case, problem (1) takes the form

$$\min \quad \|(D[c^1|\ldots|c^k])x + c\mu - b\|
\text{s.t.} \quad x \in \mathbb{R}^d, \quad \mu \in \mathbb{R}
\quad |\text{supp}(x)| \leq \sigma,$$

where $D \in \mathbb{R}^{n \times n}$ is diagonal, $b, c, c^1, \ldots, c^k \in \mathbb{R}^n$, and $\sigma \in \mathbb{N}$. The main result of this section is the following theorem.

**Theorem 1.** Problem (6) can be solved in polynomial time for varying $n$, provided that $k$ is a fixed number.

In order to prove Theorem 1 we first consider a simpler setting.

3.1 A simpler diagonal problem

In this section we consider a simpler version of problem (6), which is essentially obtained by fixing the last $k$ components of $x$ and the variable $\mu$. Formally, we consider the problem

$$\min \quad \| Dx - b \|
\text{s.t.} \quad x \in \mathbb{R}^n
\quad |\text{supp}(x)| \leq \sigma,$$

where $D \in \mathbb{R}^{n \times n}$ is diagonal, $b \in \mathbb{R}^n$, and $\sigma \in \mathbb{N}$.

**Lemma 2.** Problem (7) can be solved in polynomial time. In particular, an optimal solution is given by

$$x^*_j := \begin{cases} b_j/d_j & \text{if } j \in M \\ 0 & \text{if } j \notin M, \end{cases}$$

where $M$ is a subset of $\{1, \ldots, n\}$ of cardinality $\sigma$ with the property that for each $i \in M$ and each $j \notin M$, we have $|b_i| \geq |b_j|$.
Proof. Denote by $d_1, d_2, \ldots, d_n$ the diagonal entries of $D$. Then problem (7) is equivalent to

$$\min \sum_{j=1}^{n} (d_j x_j - b_j)^2 \quad \text{s.t.} \quad x \in \mathbb{R}^n, \quad |\text{supp}(x)| \leq \sigma,$$

(8)

The separability of the objective function implies that the objective value of a vector $x$ will be at least $\sum_{j \notin \text{supp}(x)} b_j^2$. Hence, by definition of the index set $M$, the optimum value of (8) is at least $\sum_{j \notin M} b_j^2$. The vector $x^*$ is then an optimal solution since its objective value is $\sum_{j \notin M} b_j^2$.

### 3.2 Proof of Theorem 1

By Lemma 1, in order to prove Theorem 1, we only need to show that we can solve in polynomial time a problem of the form

$$\min \left\| Dx - \left( b - \sum_{\ell=1}^{k} c_{i}^{\ell} \lambda_{\ell} \right) \right\| \quad \text{s.t.} \quad x \in \mathbb{R}^n, \quad \lambda \in \mathbb{R}^k, \quad |\text{supp}(x)| \leq \sigma,$$

(9)

provided that $k$ is a fixed number. In the remainder of the proof we show how to solve problem (9). We define a restricted version of problem (9), where we fix the variables $\lambda_{\ell}, \ell = 1, \ldots, k$,

$$\text{Opt}(\sigma)_{\lambda} := \min \left\{ \left\| Dx - \left( b - \sum_{\ell=1}^{k} c_{i}^{\ell} \lambda_{\ell} \right) \right\| : x \in \mathbb{R}^n, \quad |\text{supp}(x)| \leq \sigma \right\}.$$

**Claim 1.** We can construct in polynomial time a polynomial number of polyhedra $Q^t \subseteq \mathbb{R}^k$, for $t \in T$, that cover $\mathbb{R}^k$, and index sets $\chi^t \subseteq \{1, \ldots, n\}$ of cardinality $\sigma$, for each $t \in T$, with the following property: For each $t \in T$, the problem $\text{Opt}(\sigma)_{\lambda}$ has an optimal solution with support contained in $\chi^t$, for all $\lambda$ such that $\lambda \in Q^t$.

**Proof of claim.** Lemma 2 implies that in order to understand the optimal support of problem $\text{Opt}(\sigma)_{\lambda}$ for a fixed vector $\lambda$, it is sufficient to compare all quantities $|b_i - \sum_{\ell=1}^{k} c_{i}^{\ell} \lambda_{\ell}|$, for $i \in \{1, \ldots, n\}$. So let $i$ and $j$ be two distinct indices in $\{1, \ldots, n\}$. We wish to subdivide all points $\lambda \in \mathbb{R}^k$ based on which of the two quantities $|b_i - \sum_{\ell=1}^{k} c_{i}^{\ell} \lambda_{\ell}|$ and $|b_j - \sum_{\ell=1}^{k} c_{j}^{\ell} \lambda_{\ell}|$ is larger. In order to do so, consider the inequality

$$|b_i - \sum_{\ell=1}^{k} c_{i}^{\ell} \lambda_{\ell}| \geq |b_j - \sum_{\ell=1}^{k} c_{j}^{\ell} \lambda_{\ell}|.$$

It is simple to check that the set of points in $\mathbb{R}^k$ that satisfy the above inequality can be written as the union of polyhedra using linear inequalities corresponding to the four hyperplanes in $\mathbb{R}^k$ defined by equations

$$b_i - \sum_{\ell=1}^{k} c_{i}^{\ell} \lambda_{\ell} = 0, \quad b_j - \sum_{\ell=1}^{k} c_{i}^{\ell} \lambda_{\ell} = b_j - \sum_{\ell=1}^{k} c_{j}^{\ell} \lambda_{\ell},$$

$$b_j - \sum_{\ell=1}^{k} c_{j}^{\ell} \lambda_{\ell} = 0, \quad b_i - \sum_{\ell=1}^{k} c_{j}^{\ell} \lambda_{\ell} = - \left( b_j - \sum_{\ell=1}^{k} c_{j}^{\ell} \lambda_{\ell} \right).$$
By considering these four hyperplanes for all possible distinct pairs of indices in \( \{1, \ldots, n\} \), we obtain \( 4(n^2 - n) = O(n^2) \) hyperplanes in \( \mathbb{R}^k \). These hyperplanes subdivide \( \mathbb{R}^k \) into a number of polyhedra. By the hyperplane arrangement theorem \([6]\), this subdivision consists of at most \( O((n^2)^k) = O(n^{2k}) \) polyhedra \( Q^t \), for \( t \in T \). Since \( k \) is fixed, \( |T| \) is polynomial in \( n \) and the subdivision can be obtained in polynomial time.

We now fix one polyhedron \( Q^t \), for some \( t \in T \). By checking, for each hyperplane that we have constructed above, in which of the two half-spaces lies \( \chi^t \), we obtain a total order on all the expressions \( |b_i - \sum_{\ell=1}^{k} c^{t}_{i \ell} \lambda_{\ell}| \), for \( i \in \{1, \ldots, n\} \). The obtained total order is global, in the sense that, for each fixed \( \lambda \) with \( \lambda \in Q^t \), it induces a consistent total order on the values obtained by fixing \( \lambda \) in the expressions \( |b_i - \sum_{\ell=1}^{k} c^{t}_{i \ell} \lambda_{\ell}| \), for \( i \in \{1, \ldots, n\} \). This total order induces an ordering \( i_1, i_2, \ldots, i_n \) of the indices \( 1, \ldots, n \) such that, for every \( \lambda \in Q^t \), we have

\[
|b_{i_1} - \sum_{\ell=1}^{k} c^{t}_{i_1 \ell} \lambda_{\ell}| \geq \cdots \geq |b_{i_k} - \sum_{\ell=1}^{k} c^{t}_{i_k \ell} \lambda_{\ell}|
\]

By Lemma \([2]\), for each \( \lambda \) such that \( \lambda \in Q^t \), the problem \( \text{Opt}(\sigma)_{|\lambda} \) has an optimal solution with support contained in \( \chi^t := \{i_1, i_2, \ldots, i_n\} \). \(\diamondsuit\)

Let \( \mathcal{X} \) be the set containing all index sets \( \chi^t \) obtained in Claim \([1]\) namely

\[
\mathcal{X} := \{\chi^t : t \in T\}.
\]

**Claim 2.** There exists an optimal solution \( (x^*, \lambda^*) \) of problem \([9]\) such that

\[
\text{supp}(x^*) \subseteq \chi \text{ for some } \chi \in \mathcal{X}.
\]

**Proof of claim.** Let \( (x^*, \lambda^*) \) be an optimal solution of problem \([9]\). Then \( x^* \) is an optimal solution of the restricted problem \( \text{Opt}(\sigma)_{|\lambda^*} \). Let \( Q^t \), for \( t \in T \), be a polyhedron such that \( \lambda^* \in Q^t \), and let \( \chi^t \in \mathcal{X} \) be the corresponding index set. By Claim \([1]\) the problem \( \text{Opt}(\sigma)_{|\lambda^*} \) has an optimal solution \( \tilde{x} \) with support contained in \( \chi^t \). This implies that the solution \( (\tilde{x}, \lambda^*) \) is also optimal for problem \([9]\). \(\diamondsuit\)

For each \( \chi \in \mathcal{X} \), each problem \([9]\), with the additional constraints \( x_i = 0 \), for all \( i \notin \chi \), can then be solved in polynomial time since the cardinality constraint can be dropped, and the objective function is convex. The best solution among the obtained ones is an optimal solution of \([9]\). This concludes the proof of Theorem \([1]\). \(\square\)

Note that the algorithm for the simpler case given in Lemma \([2]\) plays a key role in the proof of Theorem \([1]\). In fact, in Theorem \([1]\) we are able to subdivide the space of the \( \lambda \) variables in a polynomial number of regions such that in each region the algorithm given in Lemma \([2]\) yields the same optimal support.

### 4 The block diagonal case

In this section we consider problem \([1]\), where the matrix \( M \) is of the general form \([2]\). In this case we have seen that problem \([1]\) takes the form \([3]\), where \( A \in \mathbb{R}^{m \times n} \) is block diagonal with blocks \( A_i \in \mathbb{R}^{m_i \times n_i} \), for \( i = 1, \ldots, h \), where \( b, c_1, \ldots, c_k \in \mathbb{R}^{m_i} \), and \( \sigma \in \mathbb{N} \). The main result of this section is the following theorem.

**Theorem 2.** Problem \([3]\) can be solved in polynomial time for varying \( n \), provided that \( k, n_1, \ldots, n_h \) are fixed numbers.
Our overall strategy to prove Theorem 2 is similar to the one we used to prove Theorem 1. Namely, we first design a polynomial-time algorithm to solve the simpler problem obtained from (3) by fixing the last \( k \) components of \( x \) and the variable \( \mu \), and by squaring the objective function, namely

\[
\begin{align*}
\min \quad & \|Ax - b\|^2 \\
\text{s.t.} \quad & x \in \mathbb{R}^n \\
& |\text{supp}(x)| \leq \sigma,
\end{align*}
\]

where \( A \in \mathbb{R}^{m \times n} \) is block diagonal, \( b \in \mathbb{R}^m \), and \( \sigma \in \mathbb{N} \). Then, we cover the space of the \( \lambda \) variables with a polynomial number of regions such that in each region the algorithm yields the same optimal support.

Note that there are many possible polynomial-time algorithms that one can devise for problem (10), and a particularly elegant one can be obtained with a dynamic programming approach similar to the classic dynamic programming recursion for knapsack [15]. For each \( i = 1, \ldots, h \), let \( x^i \in \mathbb{R}^{n^i} \) and \( b^i \in \mathbb{R}^{m^i} \) such that

\[
x = \begin{pmatrix} x^1 \\ x^2 \\ \vdots \\ x^h \end{pmatrix}, \quad b = \begin{pmatrix} b^1 \\ b^2 \\ \vdots \\ b^h \end{pmatrix}.
\]

Consider the subproblem of (10) on blocks \( A^1, \ldots, A^i \), with \( i \in \{1, \ldots, h\} \) and support \( j \), with \( j \in \{0, \ldots, \sigma\} \),

\[
\min \left\{ \left\| \begin{pmatrix} A^1 & A^2 & \cdots & A^i \\ x^1 & x^2 & \cdots & x^h \end{pmatrix} \right\|^2 : x \in \mathbb{R}^{n_1 + \cdots + n^i}, \ |\text{supp}(x)| \leq j \right\},
\]

and denote it by \( \text{Opt}(1, \ldots, i; j) \). By exploiting the separability of the objective function, it can be checked that the recursion

\[
\text{Opt}(1, \ldots, i; j) = \min_{t=0, \ldots, j} \{ \text{Opt}(1, \ldots, i-1; j-t) + \text{Opt}(i; t) \}, \quad \forall j \in \{0, \ldots, \sigma\},
\]

for \( i = 1, \ldots, h \), yields an optimal solution for problem (10).

We view this dynamic programming recursion as a “horizontal approach” where we fix the support but then proceed a recursion over the blocks. We did not see how this algorithm can be extended in order to cope with the complication arising from the presence of an extra number of columns. In order to tackle also the case with a constant number of extra columns we need a sort of “vertical approach” where we incorporate all blocks simultaneously and then develop a recursion to enlarge the support condition. In the next section, we present a combinatorial algorithm that not only can solve problem (10), but that also allows us to obtain a covering of the space of the \( \lambda \) variables that consists of a polynomial number of regions such that in each region the new algorithm yields the same optimal support. The algorithm is based on a proximity theorem between optimal solutions w.r.t. two consecutive support values.
4.1 A proximity theorem

For every $i \in \{1, \ldots, h\}$, and every $j \in \{0, \ldots, n_i\}$, let $\text{Opt}(i; j)$ be a real number. For $\sigma \in \{0, \ldots, \sum_{i=1}^{h} n_i\}$, we define

$$\text{Opt}(\sigma) := \min \left\{ \sum_{i=1}^{h} \text{Opt}(i; j_i) : \sum_{i=1}^{h} j_i = \sigma, j_i \in \{0, \ldots, n_i\} \right\}. \tag{11}$$

Note that problem (10) can be polynomially transformed to a problem $\text{Opt}(\sigma)$. This can be seen by exploiting the block diagonal structure of the matrix $A$, and by defining $\text{Opt}(i; j_i)$ to be the optimal value of the problem restricted to block $i$ and support $j_i$. The details of this reduction, albeit with a slightly different notation, are given in the proof of Theorem 2.

**Definition 1.** Let $s$ and $q$ be nonnegative integers. Given a feasible solution $j^s = (j_1^s, \ldots, j_h^s)$ for $\text{Opt}(s)$, we say that a feasible solution $j^{s+1} = (j_1^{s+1}, \ldots, j_h^{s+1})$ for $\text{Opt}(s+1)$ is $q$-close to $j^s$ if

$$\sum_{i \in I^-} (j_i^s - j_i^{s+1}) = q,$$

where $I^- := \{i \in \{1, \ldots, h\} : j_i^{s+1} < j_i^s\}$.

Clearly, if $j^{s+1}$ is $q$-close to $j^s$, we also have that

$$\sum_{i \in I^+} (j_i^{s+1} - j_i^s) = q + 1,$$

where $I^+ := \{i \in \{1, \ldots, h\} : j_i^{s+1} > j_i^s\}$. This yields to a $l_1$-proximity bound of $\|j^{s+1} - j^s\|_1 \leq 2q + 1$.

A weak composition of an integer $q$ into $p$ parts is a sequence of $p$ non-negative integers that sum up to $q$. Two sequences that differ in the order of their terms define different weak compositions. It is well-known that the number of weak compositions of a number $q$ into $p$ parts is $\binom{q+p-1}{p-1}$. For more details on weak compositions see, for example, [12].

Our next result establishes that optimal solutions for $\text{Opt}(s)$ and $\text{Opt}(s+1)$ are close to each other.

**Lemma 3** (Proximity of optimal solutions). Given an optimal solution $j^s$ for $\text{Opt}(s)$, there exists an optimal solution $j^{s+1}$ for $\text{Opt}(s+1)$ that is $q$-close to $j^s$, for some $q \in \{0, \ldots, (\theta-1)\theta(\theta+1)/2\}$, where $\theta := \max\{n_i : i = 1, \ldots, h\}$.

**Proof.** Let $j^{s+1}$ be an optimal solution for $\text{Opt}(s+1)$ such that

$$\sum_{i=1}^{h} |j_i^{s+1} - j_i^s| \text{ is minimal}. \tag{12}$$

Let $I^+ := \{i \in \{1, \ldots, h\} : j_i^{s+1} > j_i^s\}$ and $I^- := \{i \in \{1, \ldots, h\} : j_i^{s+1} < j_i^s\}$. Note that

$$\sum_{i \in I^+} (j_i^{s+1} - j_i^s) = \sum_{i \in I^-} (j_i^s - j_i^{s+1}) + 1.$$

For $p = 1, \ldots, \theta$, let

$$x_p := |\{i \in I^+ : j_i^{s+1} - j_i^s = p\}|$$

$$y_p := |\{i \in I^- : j_i^s - j_i^{s+1} = p\}|,$$
This yields to the two equations

\[ \sum_{i \in I^+} (j_i^{s+1} - j_i^*) = \sum_{p=1}^\theta px_p \quad \text{and} \quad \sum_{i \in I^-} (j_i^s - j_i^{s+1}) = \sum_{p=1}^\theta py_p. \]

If \( \sum_{p=1}^\theta py_p \leq (\theta - 1)\theta(\theta + 1)/2 \), then the statement is verified. Otherwise, \( \sum_{p=1}^\theta py_p > (\theta - 1)\theta(\theta + 1)/2 \). This implies that there exists \( v \in \{1, \ldots, \theta\} \) with \( y_v \geq \theta \). To see this, note that if \( y_p \leq \theta - 1 \), for all \( p = 1, \ldots, \theta \), then we obtain

\[ \sum_{p=1}^\theta py_p \leq (\theta - 1)\sum_{p=1}^\theta p = (\theta - 1)\theta(\theta + 1)/2. \]

From the fact that \( \sum_{p=1}^\theta px_p = \sum_{p=1}^\theta py_p + 1 > (\theta - 1)\theta(\theta + 1)/2 \), it also follows that there exists \( u \in \{1, \ldots, \theta\} \) with \( x_u \geq \theta \).

In particular we have that \( x_u \geq \theta \geq v \) and \( y_v \geq u \).

Thus there exists a subset \( \tilde{I}^+ \) of \( I^+ \) such that \( |\tilde{I}^+| = v \), and \( j_i^{s+1} - j_i^* = u \) for all \( i \in \tilde{I}^+ \).

Symmetrically, there exists a subset \( \tilde{I}^- \) of \( I^- \) such that \( |\tilde{I}^-| = u \), and \( j_i^s - j_i^{s+1} = v \) for all \( i \in \tilde{I}^- \).

Let \( \tilde{j}_i^{s+1} \) be obtained from \( j_i^{s+1} \) as follows

\[ \tilde{j}_i^{s+1} = \begin{cases} \ j_i^{s+1} - u = j_i^* & \text{if } i \in \tilde{I}^+ \\ \ j_i^{s+1} + v = j_i^* & \text{if } i \in \tilde{I}^- \\ \ j_i^{s+1} & \text{otherwise.} \end{cases} \]

Since \( \sum_{i=1}^h \tilde{j}_i^{s+1} = \sum_{i=1}^h j_i^{s+1} - uv + uv = s + 1 \), we have that \( \tilde{j}_i^{s+1} \) is a feasible solution for \( \text{Opt}(s + 1) \). Moreover, we have that

\[ \sum_{i=1}^h |\tilde{j}_i^{s+1} - j_i^*| = \sum_{i=1}^h |j_i^{s+1} - j_i^*| - 2uv < \sum_{i=1}^h |j_i^{s+1} - j_i^*|. \]

In the remainder of the proof, we show that \( \tilde{j}_i^{s+1} \) is an optimal solution for \( \text{Opt}(s + 1) \). This will conclude the proof, since it contradicts the choice of \( j_i^{s+1} \) in Eq. [12].

Let \( \tilde{j}_i^s \) be obtained from \( j_i^s \) as follows

\[ \tilde{j}_i^s = \begin{cases} \ j_i^s + u = j_i^{s+1} & \text{if } i \in \tilde{I}^+ \\ \ j_i^s - v = j_i^{s+1} & \text{if } i \in \tilde{I}^- \\ \ j_i^s & \text{otherwise.} \end{cases} \]

Since \( \sum_{i=1}^h \tilde{j}_i^s = \sum_{i=1}^h j_i^s + uv - uv = s \), we have that \( \tilde{j}_i^s \) is a feasible solution for \( \text{Opt}(s) \). Since \( \tilde{j}_i^s \) is an optimal solution for \( \text{Opt}(s) \) we obtain

\[ \sum_{i=1}^h \text{Opt}(i; \tilde{j}_i^s) - \sum_{i=1}^h \text{Opt}(i; j_i^s) = \sum_{i \in \tilde{I}^+ \cup \tilde{I}^-} (\text{Opt}(i; j_i^{s+1}) - \text{Opt}(i; j_i^*)) \geq 0. \]
Consider now the objective value of the feasible solution $\tilde{j}^{s+1}$ for Opt$(s + 1)$.

$$
\sum_{i=1}^{h} \text{Opt}(i; \tilde{j}^{s+1}) = \sum_{i=1}^{h} \text{Opt}(i; j^{s+1}_i) + \sum_{i \in I^+ \cup I^-} (\text{Opt}(i; j^{s+1}_i) - \text{Opt}(i; j^{s+1})) \\
\leq \sum_{i=1}^{h} \text{Opt}(i; j^{s+1}_i).
$$

This shows that the solution $\tilde{j}^{s+1}$ is optimal for Opt$(s + 1)$. \hfill \Box

Assume now that we know an optimal solution $j^s$ for Opt$(s)$ and we wish to obtain an optimal solution for Opt$(s + 1)$. By Lemma 3, we just need to consider the feasible solutions for Opt$(s + 1)$ that are $q$-close to $j^s$, for some $q \leq (\bar{\theta} - 1)\bar{\theta}(\bar{\theta} + 1)/2 =: \bar{\theta}$. We denote the family of these solutions by Aug$(j^s)$, formally

$$
\text{Aug}(j^s) := \left\{ j^{s+1} : j^{s+1}_i \in \{0, \ldots, n_i\}, i \in \{1, \ldots, h\}, \sum_{i=1}^{h} j^{s+1}_i = s + 1, j^{s+1} \text{ is } q\text{-close to } j^s \text{ for some } q \leq \bar{\theta} \right\}. \quad (13)
$$

For each solution $j^{s+1} \in \text{Aug}(j^s)$, the corresponding objective function value is obtained from Opt$(s)$ by adding the difference,

$$
d(j^s, j^{s+1}) := \sum_{i \in I^+ \cup I^-} (\text{Opt}(i; j^{s+1}_i) - \text{Opt}(i; j^s_i)). \quad (14)
$$

We denote by $\mathcal{D}(j^s)$ the family of the values $d(j^s, j^{s+1})$, for each solution $j^{s+1} \in \text{Aug}(j^s)$,

$$
\mathcal{D}(j^s) := \{d(j^s, j^{s+1}) : j^{s+1} \in \text{Aug}(j^s)\}.
$$

From our discussions it follows that, in order to select the optimal solution for Opt$(s + 1)$, we only need to know which value in $\mathcal{D}(j^s)$ is the smallest.

We next define the set $\mathcal{D}$ as the union of all sets $\mathcal{D}(j^s)$, for any feasible solution $j^s$ of problem Opt$(s)$, for any $s \in \{0, \ldots, \sum_{i=1}^{h} n_i\}$. Formally, the set $\mathcal{D}$ is defined as

$$
\mathcal{D} := \left\{d(j^s, j^{s+1}) : j^s \text{ is a feasible solution for Opt}(s), \text{ for some } s \in \{0, \ldots, \sum_{i=1}^{h} n_i\}, j^{s+1} \in \text{Aug}(j^s) \right\}. \quad (15)
$$

The next result implies that the set $\mathcal{D}$ contains a number of values that is polynomial in $h$, provided that $\theta$ is fixed. This fact is on the first glance surprising since the number of feasible solutions $j^s$ for Opt$(s)$ is of exponential order $\theta^h$.

**Lemma 4.** The set $\mathcal{D}$ contains $O((\bar{\theta} + h)^{\bar{\theta} + 1}2^{\bar{\theta} + 1})$ values.

**Proof.** We count the number of all possible values $d(j^s, j^{s+1})$ in $\mathcal{D}$. Fix $q \in \{0, \ldots, \bar{\theta}\}$, and let us consider all the values $d(j^s, j^{s+1})$ corresponding to a feasible solution $j^s$ for Opt$(s)$ and a feasible solution $j^{s+1}$ for Opt$(s + 1)$ such that $j^{s+1}$ is $q$-close to $j^s$. First, we construct the possible
Proof of Theorem 2. We apply Lemma 1 and then square the objective function of the obtained problem of the form (16), provided that \( k, n_1, \ldots, n_k \) are fixed numbers. In particular, the set \( \lambda \) of positive differences \( d_i := j_i^{+1} - j_i^+ \), for \( i \in I^+ \) is fixed.

Given a total order on all the values in \( D \), we can construct an optimal solution for \( \text{Opt}(\sigma) \), for any \( \sigma \in \{0, \ldots, \sum_{i=1}^b n_i\} \), in time \( O(\sigma(\theta + h)^{\theta + 1q_2q_1 + 1}) \).

Proof. An optimal solution for \( \text{Opt}(0) \) is \( j^0 = (0, \ldots, 0) \). Let \( s \in \{0, \ldots, \sum_{i=1}^b n_i\} \) and assume that we have an optimal solution \( j^s \) for \( \text{Opt}(s) \). We show how we can construct an optimal solution \( j^{s+1} \) for \( \text{Opt}(s+1) \). Consider all the values in \( D(j^s) \). Since \( D(j^s) \subseteq D \), we can inquire for a total order of \( D(j^s) \) from the given total order of \( D \). Thus \( D(j^s) \) has a minimum element. Since \( \lambda \) is bounded by \( \theta(\theta + h)^{\theta + 1q_2q_1 + 1} \), we can be constructed in linear time with respect to the cardinality of the set, in our case it can be found in time \( O(\sigma(\theta + h)^{\theta + 1q_2q_1 + 1}) \) as a consequence of Lemma 4. In view of Lemma 3 the solution in \( \text{Aug}(j^s) \) corresponding to the minimum element in \( D(j^s) \) is an optimal solution \( j^{s+1} \) for \( \text{Opt}(s+1) \). This argument applied in an inductive manner leads to an optimal solution for \( \text{Opt}(\sigma) \) for any \( \sigma \in \{0, \ldots, \sum_{i=1}^b n_i\} \). \( \square \)

4.2 Proof of Theorem 2

Before proceeding to the formal proof of Theorem 2, we give a brief overview of the proof. First, in view of Lemma 1 we reduce problem 3 to a fixed number of problems of the form

\[
\min \left\| Ax - \left( b - \sum_{\ell=1}^k c_{\ell} \lambda_{\ell} \right) \right\|^2
\]

s.t. \( x \in \mathbb{R}^n \), \( \lambda \in \mathbb{R}^k \)

\( |\text{supp}(x)| \leq \sigma \),

where \( A \in \mathbb{R}^{m \times n} \) is block diagonal with blocks \( A^i \in \mathbb{R}^{m_i \times n_i} \), for \( i = 1, \ldots, h \), where \( b, c_1, \ldots, c_k \in \mathbb{R}^m \), \( \sigma \in \mathbb{N} \), and \( k \) is fixed.

The rest of the proof is dedicated to the solution of (16). For each fixed \( \lambda \in \mathbb{R}^k \), problem (16) admits an optimal solution with a specific support of the optimal \( x \) vector. Our aim is to partition all possible values of \( \lambda \) based on the optimal support that they yield. This is verified in two steps. Claim 3 and Claim 4 establish the details. In view of these two claims we can conclude that there is a set \( \chi \) that contains, for each element in the partition, the corresponding optimal support. In particular, the set \( \chi \) has the property that there is at least one optimal solution of (16) whose support is contained in some \( \chi \in \mathcal{X} \). Finally, for each support \( \chi \in \mathcal{X} \), an optimal solution of problem (16) with support contained in \( \chi \) can be solved in polynomial time.

We now give the formal proof of Theorem 2.

Proof of Theorem 2. We apply Lemma 1 and then square the objective function of the obtained problems. Hence, in order to prove Theorem 2, we only need to show that we can solve in polynomial time a problem of the form (16), provided that \( k, n_1, \ldots, n_h \) are fixed numbers. In particular, the set \( \lambda \) of positive differences \( d_i := j_i^{+1} - j_i^+ \), for \( i \in I^+ \) is fixed.
the remainder of the proof we show how to solve problem (16). We define a restricted version of problem (16), where we fix the variables \( \lambda_\ell, \ell = 1, \ldots, k \),

\[
\text{Opt}(\sigma|_\lambda) := \min \left\{ \left\| Ax - \left( b - \sum_{\ell=1}^{k} c_\ell \lambda_\ell \right) \right\|^2 : x \in \mathbb{R}^n, \ |\text{supp}(x)| \leq \sigma \right\}.
\]

We now wish to rewrite \( \text{Opt}(\sigma|_\lambda) \) by exploiting the separability of the objective function. For each \( i = 1, \ldots, h \), let \( x^i \in \mathbb{R}^{n_i} \) and \( b^i \in \mathbb{R}^{m_i} \) such that

\[
x = \begin{pmatrix} x^1 \\ x^2 \\ \vdots \\ x^h \end{pmatrix}, \quad b = \begin{pmatrix} b^1 \\ b^2 \\ \vdots \\ b^h \end{pmatrix}.
\]

For each \( i = 1, \ldots, h \), and each \( \ell = 1, \ldots, k \), let \( c_\ell^i \in \mathbb{R}^{m_i} \) such that

\[
c_\ell = \begin{pmatrix} c_\ell^1 \\ c_\ell^2 \\ \vdots \\ c_\ell^h \end{pmatrix}.
\]

Consider the subproblem of \( \text{Opt}(\sigma|_\lambda) \) on block \( A^i \), with \( i \in \{1, \ldots, h\} \), and support \( j \), for \( j \in \{0, \ldots, n_i\} \). Formally,

\[
\text{Opt}(i; j|_\lambda) := \min \left\{ \left\| A^i x^i - \left( b^i - \sum_{\ell=1}^{k} c_\ell^i \lambda_\ell \right) \right\|^2 : x^i \in \mathbb{R}^{n_i}, \ |\text{supp}(x^i)| \leq j \right\},
\]

We can finally rewrite \( \text{Opt}(\sigma|_\lambda) \) in the form

\[
\text{Opt}(\sigma|_\lambda) = \min \left\{ \sum_{i=1}^{h} \text{Opt}(i; j_i|_\lambda) : \sum_{i=1}^{h} j_i = \sigma, \ j_i \in \{0, \ldots, n_i\} \right\}.
\]

(17)

Note that in this new form, the decision variables are the integers \( j_i \), for \( i = 1, \ldots, h \), and the variables \( x \) do not appear explicitly. We observe that we have reduced ourselves to the same setting described in Section 4.1. In fact, for each fixed \( \lambda \), each \( \text{Opt}(i; j_i|_\lambda) \) can be calculated in polynomial time. Thus, problem \( \text{Opt}(\sigma|_\lambda) \) is now a problem of the form \( \text{Opt}(\sigma) \), as defined in (11) and can be solved efficiently as a consequence of Proposition 1. Hence, problem \( \text{Opt}(\sigma|_\lambda) \) can be solved for each fixed \( \lambda \). However, in order to solve our original problem, we have to solve \( \text{Opt}(\sigma|_\lambda) \) for every \( \lambda \in \mathbb{R}^k \). In order to do this, we now think of \( \text{Opt}(\sigma|_\lambda) \) and \( \text{Opt}(i; j_i|_\lambda) \) as functions that associate to each \( \lambda \in \mathbb{R}^k \) a real number.

Next, we define a space \( S \) that is an extended version of the space \( \mathbb{R}^k \) of variables \( \lambda_\ell \), for \( \ell = 1, \ldots, k \). The space \( S \) contains all the variables \( \lambda_\ell \), for \( \ell = 1, \ldots, k \), and it also contains one variable for each product of two variables \( \lambda_{\ell_1}\lambda_{\ell_2} \), with \( \ell_1, \ell_2 \in \{1, \ldots, k\} \). The dimension of the space \( S \) is therefore \( O(k^2) \). Note that, for each \( \lambda \in \mathbb{R}^k \), there exists a unique corresponding point in \( S \), that we denote by \( \text{ext}(\lambda) \), obtained by computing all the products \( \lambda_{\ell_1}\lambda_{\ell_2} \), for \( \ell_1, \ell_2 \in \{1, \ldots, k\} \).

Claim 3. We can construct in polynomial time a polynomial number of polyhedra \( P^i \subseteq S \), for \( i \in T \), that cover \( S \), and index sets \( v^i(i; j) \subseteq \{1, \ldots, n_i\} \) of cardinality \( j \), for each \( i \in \{1, \ldots, h\} \),
Let $t \in T$. We can construct in polynomial time a polynomial number of polyhedra
$Q^{t,u} \subseteq P^t$, for $u \in U^t$, that cover $P^t$, and index sets $\chi^{t,u} \subseteq \{1, \ldots, n\}$ of cardinality $\sigma$, for each $u \in U^t$, with the following property: The problem Opt($\sigma$)$_{\ell}$ has an optimal solution with support contained in $\chi^{t,u}$, for all $\lambda$ such that ext($\lambda$) $\in Q^{t,u}$.
Proof of claim. To prove this claim, we first construct all polyhedra \(Q^t,u\), for \(u \in U^t\). Then we show how to construct the index sets with the desired property.

Let \(s \in \{0, \ldots, n\}\). Let \(j^s = (j^s_1, \ldots, j^s_h)\) such that \(j^s_i \in \{0, \ldots, n_i\}\) for every \(i \in \{1, \ldots, h\}\), and \(\sum_{i=1}^h j^s_i = s\). As in Section 4.1 let \(\theta := \max\{n_i : i = 1, \ldots, h\}\), \(\bar{\theta} := (\theta - 1)\theta(\theta + 1)/2\), and define \(P^s\) as in [13]. Let \(j^{s+1} \in \Lambda\). Define the sets \(I^- := \{i \in \{1, \ldots, h\} : j^{s+1}_i < j^s_i\}\) and \(I^+ := \{i \in \{1, \ldots, h\} : j^{s+1}_i > j^s_i\}\). For \(\lambda\) such that \(\text{ext}(\lambda) \in P^t\), we define the expression

\[
d(j^s, j^{s+1})_{\lambda} := \sum_{i \in I^+ \cup I^-} (\text{Opt}(i; j^{s+1})_{\lambda} - \text{Opt}(i; j^s)_{\lambda}).
\]

For \(\lambda\) such that \(\text{ext}(\lambda) \in P^t\), consider the set \(D_{\lambda}\) defined by

\[
D_{\lambda} := \left\{d(j^s, j^{s+1})_{\lambda} : \text{there exists } s \in \{0, \ldots, \sum_{i=1}^h n_i\} \text{ such that } j^s \text{ is a feasible solution for } \text{Opt}(s)_{\lambda} \text{ and } j^{s+1} \in \text{Aug}(j^s)\right\}.
\]

Note that for each fixed \(\lambda\), each \(d(j^s, j^{s+1})_{\lambda}\) is a value of the form \(d(j^s, j^{s+1})\), as defined in (14), and the set \(D_{\lambda}\) reduces to the set \(D\), as defined in (15).

Let \(d(j^s, j^{s+1})_{\lambda}\) and \(d(k^s, k^{s+1})_{\lambda}\) be two distinct expressions in \(D_{\lambda}\). We wish to subdivide all points \(\text{ext}(\lambda) \in S\) based on which of the two expressions is larger. In order to do so, consider the equation

\[
d(j^s, j^{s+1})_{\lambda} = d(k^s, k^{s+1})_{\lambda}.
\]

We show that (19) is a quadratic equation in \(\lambda\). Consider a single \(\text{Opt}(i; j)_{\lambda}\) that appears in the expression defining \(d(j^s, j^{s+1})_{\lambda}\), and let \(u^f(i; j)\) be the corresponding index set from Claim [3]. Let \(L\) be the linear subspace of \(\mathbb{R}^{m_i}\) defined by \(L := \{A^x : \text{supp}(x') \subseteq u^f(i; j)\}\), and let \(p^f(\lambda) := b^t - \sum_{\ell=1}^k c^t_{\lambda, i}\). From Claim [3] for all \(\lambda\) such that \(\text{ext}(\lambda) \in P^t\), we have that \(\text{Opt}(i; j)_{\lambda}\) can be written as the quadratic function in \(\lambda\) of the form (18), namely

\[
\text{Opt}(i; j)_{\lambda} = \left(\text{proj}_L(p^f(\lambda)) - p^f(\lambda)\right)^2.
\]

The expression \(d(j^s, j^{s+1})_{\lambda}\) is a linear combination of expressions \(\text{Opt}(i; j)_{\lambda}\). Hence, it can also be written as a quadratic function in \(\lambda\). The same argument shows that also the expression \(d(k^s, k^{s+1})_{\lambda}\) can be written as a quadratic function in \(\lambda\). Hence (19) is a quadratic equation in \(\lambda\). By linearizing all the quadratic terms, we obtain a hyperplane in the space \(S\).

As a consequence of Lemma [3], the set \(D_{\lambda}\) contains \(O(h^{2(\theta + 1)})\) expressions. Thus, by considering the corresponding hyperplane for all possible distinct pairs of expressions in \(D_{\lambda}\), we obtain a total number of \(O(h^{2(\theta + 1)})\) hyperplanes. These hyperplanes subdivide \(S\) into a number of polyhedra. The hyperplane arrangement theorem [6] implies that this subdivision consists of at most \(O((h^{2(\theta + 1)})S) = O((h^{2(\theta + 1)})k^2)\) polyhedra \(R^u\), for \(u \in U^t\). Since \(k\) and \(\theta\) are fixed, \(|U^t|\) is polynomial in \(h\) and the subdivision can be obtained in polynomial time. Define \(Q^t,u := P^t \cap R^u\), for every \(u \in U^t\).

We now fix one polyhedron \(Q^t,u\), for some \(u \in U^t\). By checking, for each hyperplane that we have constructed above, in which of the two half-spaces lies \(Q^t,u\), we obtain a total order on all the expressions in \(D_{\lambda}\). The obtained total order is global, in the sense that, for each fixed \(\lambda\) with \(\text{ext}(\lambda) \in Q^t,u\), it induces a consistent total order on the values obtained by fixing \(\lambda\) in the expressions in \(D_{\lambda}\). Since problem \(\text{Opt}(\sigma)_{\lambda}\) can be written in the form (17), Proposition [1] implies that we can obtain an optimal support \(\chi^t,u \subseteq \{1, \ldots, \sum_{i=1}^h n_i\}\) for problem \(\text{Opt}(\sigma)_{\lambda}\), for each fixed \(\lambda\) with \(\text{ext}(\lambda) \in Q^t,u\). Note that, since the total order is independent on \(\lambda\), also the obtained support is independent on \(\lambda\). Therefore the claim follows. \(\diamondsuit\)
Let $\mathcal{X}$ be the set containing all index sets $\chi^{t,u}$ obtained in Claim 4, namely

$$\mathcal{X} := \{\chi^{t,u} : t \in T, u \in U^t\}.$$ 

**Claim 5.** There exists an optimal solution $(x^*, \lambda^*)$ of problem (16) such that $\text{supp}(x^*) \subseteq \chi$ for some $\chi \in \mathcal{X}$.

**Proof of claim.** Let $(x^*, \lambda^*)$ be an optimal solution of problem (16). Then $x^*$ is an optimal solution of the restricted problem $\text{Opt}(\sigma)|_{\lambda^*}$. Let $Q^{t,u}$, for $t \in T$, $u \in U^t$, be a polyhedron such that $\text{ext}(\lambda^*) \in Q^{t,u}$, and let $\chi^{t,u} \in \mathcal{X}$ be the corresponding index set. From Claim 4, the problem $\text{Opt}(\sigma)|_{\lambda^*}$ has an optimal solution $\tilde{x}$ with support contained in $\chi^{t,u}$. This implies that the solution $(\tilde{x}, \lambda^*)$ is also optimal for problem (16). \hfill \Box

For each $\chi \in \mathcal{X}$, each problem (16), with the additional constraints $x_i = 0$, for all $i \notin \chi$, can then be solved in polynomial time since the cardinality constraint can be dropped, and the objective function is convex. The best solution among the obtained ones is an optimal solution of (16). This concludes the proof of Theorem 2. \hfill \Box

**References**


