FULL LENGTH PAPER

Supermodularity and valid inequalities for quadratic optimization with indicators

Alper Atamtürk[1] · **Andrés Gómez[2]**

Abstract

We study the minimization of a rank-one quadratic with indicators and show that the underlying set function obtained by projecting out the continuous variables is supermodular. Although supermodular minimization is, in general, difficult, the specific set function for the rank-one quadratic can be minimized in linear time. We show that the convex hull of the epigraph of the quadratic can be obtained from inequalities for the underlying supermodular set function by lifting them into nonlinear inequalities in the original space of variables. Explicit forms of the convex-hull description are given, both in the original space of variables and in an extended formulation via conic quadratic-representable inequalities, along with a polynomial separation algorithm. Computational experiments indicate that the lifted supermodular inequalities in conic quadratic form are quite effective in reducing the integrality gap for quadratic optimization with indicators.

✉ Alper Atamtürk
atamturk@berkeley.edu

Andrés Gómez
gomezand@usc.edu

[1] Department of Industrial Engineering and Operations Research, University of California Berkeley, Berkeley, CA 94720, USA

[2] Department of Industrial and Systems Engineering, Viterbi School of Engineering, University of Southern California, Los Angeles, CA 90089, USA

1 Introduction

Consider the convex quadratic optimization problem with indicators

$$\min\bigl\{a'x+b'y+y'Qy:\ y_i(1-x_i)=0, i=1,\ldots,n;\ (x,y)\in\{0,1\}^n\times\mathbb{R}_+^n\bigr\} \quad (1)$$

where $a,b\in\mathbb{R}^n$ and $Q\in\mathbb{R}^{n\times n}$ is a symmetric positive semi-definite matrix. For each $i=1,\ldots,n$, the binary variable x_i, along with the complementarity constraint $y_i(1-x_i)=0$, indicates whether y_i may take positive values. Problem (1) arises in numerous practical applications, including portfolio optimization [16], signal/image denoising [13, 14], best subset selection [15, 20, 34], and unit commitment [25].

Constructing strong convex relaxations for non-convex optimization problems is critical in devising effective solution approaches for them. Natural convex relaxations of (1), where the complementarity constraints $y_i(1-x_i)=0$ are linearized using the so-called "big-M" constraints $y_i\le Mx_i$, are known to be weak [40, e.g.,]. Therefore, there is an increasing effort in the literature to better understand and describe the epigraph of quadratic functions with indicator variables. Dong and Linderoth [21] describe lifted linear inequalities for (1) from its continuous quadratic optimization counterpart over bounded variables. Bienstock and Michalka [17] give a characterization of linear inequalities obtained by strengthening gradient inequalities of a convex objective function over a non-convex set.

The majority of the work toward constructing strong relaxations of (1) is based on the *perspective reformulation* [2, 18, 23, 32, 35, 39, 55, 57]. The perspective reformulation, which may be seen as a consequence of the convexifications based on disjunctive programming derived in [19], is based on strengthening the epigraph of a univariate convex quadratic function $y_i^2\le t$ by using its perspective $y_i^2/x_i\le t$. The perspective strengthening can be applied to a general convex quadratic $y'Qy$, by writing it as $y'(Q-D)y+y'Dy$ for a diagonal matrix $D\succ0$ and $Q-D\succeq0$, and simply reformulating each separable quadratic term $D_{ii}y_i^2$ as $D_{ii}y_i^2/x_i$ [22, 24, 61]. While this approach is effective when Q is strongly diagonal dominant, it is ineffective otherwise, or inapplicable when Q is not full-rank as no such D exists.

To address the limitations of the perspective reformulation, a recent stream of research focuses on constructing strong relaxations of the epigraphs of simple but multi-variable quadratic functions. Jeon et al. [36] use linear lifting to construct valid inequalities for the epigraphs of two-variable quadratic functions. Frangioni et al. [26] use extended formulations based on disjunctive programming to derive stronger relaxations of the epigraph of two-variable functions. They study heuristics and semi-definite programming (SDP) approaches to extract from Q such two-variable terms. The disjunctive approach results in a substantial increase in the size of the formulations, which limits its use to small instances. Atamtürk and Gómez [6] describe the convex hull of the epigraph of the two-variable quadratic function $(y_1-y_2)^2\le t$ in the original space of variables, and Atamtürk et al. [13] generalize this result to convex two-variable quadratic functions $a_1y_1^2-2y_1y_2+a_2y_2^2\le t$ and show how to optimally decompose an M-matrix (psd with non-positive off-diagonals) Q into such two-variable terms; their numerical results indicate that such formulations considerably improve the convex relaxations when Q is an M-matrix, but the relaxation quality degrades when Q

has positive off-diagonal entries. Han et al. [33] give SDP formulations for (1) based on convex-hull descriptions of the $2x2$ case. These SDP formulations require $O(n^2)$ additional variables and constraints, which may not scale to large problems. Wei et al. [53] give an extended formulation via a single SDP constraint and linear inequalities. Atamtürk and Gómez [7] give the convex hull description of a rank-one function with free continuous variables, and propose an SDP formulation to tackle quadratic optimization problems with free variables arising in sparse regression. Wei et al. [51, 52] extend those results, deriving ideal formulations for rank-one functions with *arbitrary* constraints on the indicator variables x. These formulations are shown to be effective in sparse regression problems; however as they do not account for the non-negativity constraints on the continuous variables, they are weak for (1). The rank-one quadratic set studied in this paper addresses this gap and properly generalizes the perspective strengthening of a univariate quadratic to higher dimensions.

In the context of discrete optimization, submodularity/supermodularity plays a critical role in the design of algorithms [27, 31, 44] and in constructing convex relaxations to discrete problems [1, 5, 10, 42, 48, 56, 58–60]. Exploiting submodularity in settings involving continuous variables as well typically require specialized arguments, e.g., see [12, 37, 49]. A notable exception is Wolsey [54], presenting a systematic approach for exploiting submodularity in fixed-charge network problems. As submodularity arises in combinatorial optimization, where the convex hulls of the sets under study are polyhedral, there are few papers utilizing submodularity to describe non-polyhedral convex hulls [8], and those sets typically involve some degree of separability between continuous and discrete variables. In this paper, we show how to generalize the valid inequalities proposed in [54] to convexify non-polyhedral sets, where the continuous variables are linked with the binary variables via indicator constraints.

Contributions

Here, we study the mixed-integer epigraph of a rank-one quadratic function with indicator variables and non-negative continuous variables:

$$X = \left\{ (x, y, t) \in \{0, 1\}^N \times \mathbb{R}_+^N \times \mathbb{R}_+ : \left(\sum_{i \in N^+} y_i - \sum_{i \in N^-} y_i \right)^2 \leq t, \; y_i(1 - x_i) = 0, \; i \in N \right\},$$

where (N^+, N^-) is a partition of $N := \{1, \ldots, n\}$. Observe that any rank-one quadratic of the form $(c'y)^2 \leq t$ with $c_i \neq 0$ for all $i \in N$ can be written as in X by scaling the continuous variables. If all coefficients of c are of the same sign, then either $N^+ = \emptyset$ or $N^- = \emptyset$, and X reduces to the simpler form

$$X_+ = \left\{ (x, y, t) \in \{0, 1\}^N \times \mathbb{R}_+^N \times \mathbb{R}_+ : \left(\sum_{i \in N} y_i \right)^2 \leq t, \; y_i(1 - x_i) = 0, \; i \in N \right\}.$$

To the best of our knowledge, the convex hull structure of X or X_+ has not been studied before. Interestingly, optimization of a linear function over X can be done in linear time (Sect. 4.2).

Our motivation for studying X stems from constructing strong convex relaxations for problem (1) by writing the convex quadratic $y'Qy$ as a sum of rank-one quadratics. Especially in large-scale applications, it is effective to state Q as a sum of a low-rank matrix and a diagonal matrix. Specifically, suppose that $Q = FF' + D$, where $F \in \mathbb{R}^{n \times r}$ and $D \in \mathbb{R}^{n \times n}$ is a (possibly equal to zero) nonnegative diagonal matrix. Such decompositions can be constructed in numerous ways, including singular-value decomposition, Cholesky decomposition, or via factor models. Letting F_j denote the j-th column of F, adding auxiliary variables $t \in \mathbb{R}^r$, $j = 1, \ldots, r$, and using the perspective reformulation, problem (1) can be cast as

$$\min_{x,y,t} a'x + b'y + \sum_{j=1}^{r} t_j + \sum_{i=1}^{n} D_{ii} \frac{y_i^2}{x_i} \tag{2a}$$

$$\text{s.t. } (F_j'y)^2 \leq t_j, \;\; j = 1, \ldots, r \tag{2b}$$

$$(x, y) \in \{0, 1\}^N \times \mathbb{R}_+^n, \; t \in \mathbb{R}^r. \tag{2c}$$

Formulation (2) arises naturally, for example, in portfolio risk minimization [16], where the covariance matrix Q is the sum of a low-rank factor covariance matrix and an idiosyncratic (diagonal) variance matrix. When the entries of the diagonal matrix D are small, the perspective reformulation is not effective in strengthening the formulation. However, noting that $(x, F_j \circ y, t_j) \in X$, where $(F_j \circ y)_i = F_{ij}y_i$, for each $j = 1, \ldots, r$, one can employ strong relaxations based on the rank-one quadratic with indicators. Our approach for decomposing $y'Qy$ into a sum of rank-one quadratics and utilizing strong relaxations of epigraphs of rank-one quadratics is analogous to employing cuts separately from individual rows of a constraint matrix $Ax \leq b$ in mixed-integer linear programming.

In this paper, we present a generic framework for obtaining valid inequalities for mixed-integer nonlinear optimization problems by exploiting supermodularity of the underlying set function. To do so, we project out the continuous variables and derive valid inequalities for the corresponding pure integer set and then lift these inequalities to the space of continuous variables as in Nguyen et al. [43], Richard and Tawarmalani [47]. It turns out that for the rank-one quadratic with indicators, the corresponding set function is supermodular and holds much of the structure of X. The *lifted supermodular inequalities* derived in this paper are *nonlinear* in both the continuous and discrete variables.

We show that this approach encompasses several previously known convexifications for quadratic optimization with indicator variables. Moreover, the well-known inequalities in the mixed-integer *linear* optimization literature given in [54], which include flow cover inequalities as a special case, can also be obtained via the lifted supermodular inequalities.

Finally, and more importantly, we show that the lifted supermodular inequalities and bound constraints are sufficient to describe cl conv(X). Such convex hull descrip-

tions of high-dimensional nonlinear sets are rare in the literature. In particular, we give a characterization in the original space of variables. This description is defined by a piecewise valid function with exponentially many pieces; therefore, it cannot be used by the convex optimization solvers directly. To overcome this difficulty, we also give a conic quadratic representable description in an extended space, with exponentially many valid conic quadratic inequalities, along with a polynomial-time separation algorithm.

The rank-one quadratic sets X and X_+ appear very similar to their relaxation

$$
X_f = \left\{ (x, y, t) \in \{0, 1\}^N \times \mathbb{R}^N \times \mathbb{R} : \left(\sum_{i \in N} y_i \right)^2 \leq t, \ y_i(1 - x_i) = 0, \ i \in N \right\},
$$

where the non-negativity constraints on the continuous variables $y \geq 0$ are dropped. However, while only one additional inequality $\frac{(\sum_{i \in N} y_i)^2}{\sum_{i \in N} x_i} \leq t$ is needed to describe cl conv(X_f) [7], the convex hulls of X and X_+ are substantially more complicated and rich. Indeed, cl conv(X_f) provides a weak relaxation for cl conv(X_+), as illustrated in the next example.

Example 1 Consider set X_+ with $n = 3$. For the relaxation X_f, the closure of the convex hull is described by $0 \leq x \leq 1$ and inequality $t \geq \frac{(y_1+y_2+y_3)^2}{\min\{1, x_1+x_2+x_3\}}$. Figure 1a depicts this inequality as a function of (x_1, y_1) for $x_2 = 0.6$, $x_3 = 0.3$, $y_2 = 0.5$, and $y_3 = 0.2$ (fixed). In Proposition 8, we give the function f describing cl conv(X_+). Figure 1b depicts $f(x, y)$ (truncated at 5) as a function of (x_1, y_1) when other variables are fixed as before.

We find that cl conv(X_f) is a very weak relaxation of cl conv(X_+) for low values of x_1. For example, for $x_1 = 0.01$ and $y_1 = 1$, we find that $\frac{(1+0.5+0.2)^2}{0.01+0.6+0.3} \approx 3.18$, whereas $f(x, y) \approx 100.55$. The computation of f for this example is described after Proposition 8. □

Outline

The rest of the paper is organized as follows. In Sect. 2 we review the valid inequalities for supermodular set functions and present the general form of the lifted supermodular inequalities. In Sect. 3 we re-derive known ideal formulations in the literature for quadratic optimization using the lifted supermodular inequalities. In Sect. 4 we show that the lifted supermodular inequalities are sufficient to describe the convex hull of X. In Sect. 5 we provide the explicit form of the lifted supermodular inequalities for X, both in the original space of variables and in a conic quadratic representable form in an extended space, and discuss the separation problem. In Sect. 6 we present computational results, and in Sect. 7 we conclude the paper.
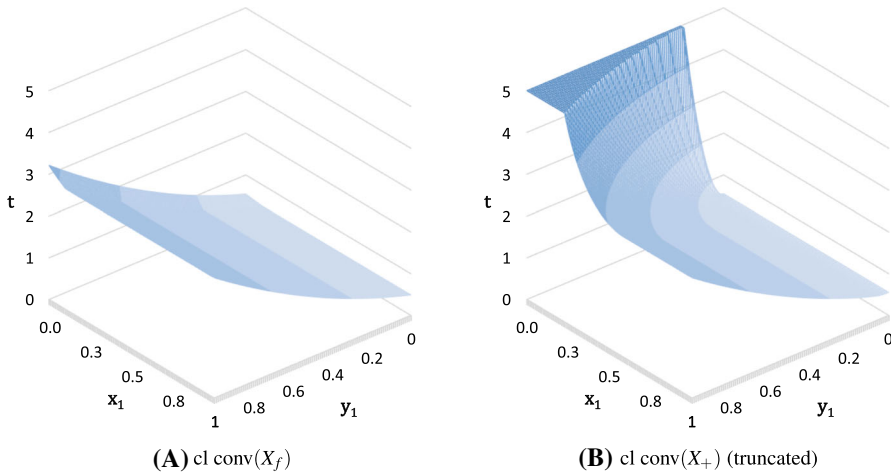
(A) cl conv(X_f) **(B)** cl conv(X_+) (truncated)

Fig. 1 Comparison of cl conv(X_f) and cl conv(X_+). Variables $x_2 = 0.6$, $x_3 = 0.3$, $y_2 = 0.5$, and $y_3 = 0.2$ are fixed

Notation

For a set $S \subseteq N$, define x_S as the indicator vector of S. By abusing notation, given a set function $g : 2^N \to \mathbb{R}$, we may equivalently write $g(S)$ or $g(x_S)$. To simplify the notation, given $i \in N$ and $S \subseteq N$, we write $S \cup i$ instead of $S \cup \{i\}$ and $S\backslash i$ instead of $S\backslash\{i\}$. For a set $Y \subseteq \mathbb{R}^N$, let conv(Y) denote the convex hull of Y and cl conv(Y) denote its closure. We adopt the convention that $a/0 = \infty$ if $a > 0$ and $a/0 = 0$ if $a = 0$. For a $a \in \mathbb{R}$, let $a_+ = \max\{a, 0\}$. For a vector $c \in \mathbb{R}^N$ and a set $S \subseteq N$, we let $c(S) = \sum_{i \in S} c_i$, $\max_c(S) = \max_{i \in S} c_i$ (by convention, $\max_c(\emptyset) = 0$) and c_S be the subvector of c induced by S. For an optimization problem with variables x, an optimal solution is denoted by x^*.

2 Preliminaries

In this section we cover a few preliminary results for the paper and, at the end, give the general form of the lifted supermodular inequalities (Theorem 1).

2.1 Supermodularity and valid inequalities

A set function $g : 2^N \to \mathbb{R}$ is *supermodular* if

$$\rho(i, S) \leq \rho(i, T) \quad \forall i \in N \text{ and } \forall S \subseteq T \subseteq N\backslash i,$$

where $\rho(i, S) = g(S \cup i) - g(S)$ is the increment function.

Proposition 1 (Nemhauser et al. [42]) *If g is a supermodular function, then*

(1) $g(T) \geq g(S) + \sum\limits_{i \in T \setminus S} \rho(i, S) - \sum\limits_{i \in S \setminus T} \rho(i, N \setminus i)$ *for all* $S, T \subseteq N$

(2) $g(T) \geq g(S) + \sum\limits_{i \in T \setminus S} \rho(i, \emptyset) - \sum\limits_{i \in S \setminus T} \rho(i, S \setminus i)$ *for all* $S, T \subseteq N$.

As a direct consequence of Proposition 1, one can construct valid inequalities for the epigraph of a supermodular function g, i.e.,

$$Z = \left\{ (x, t) \in \{0, 1\}^N \times \mathbb{R} : g(x) \leq t \right\}.$$

Specifically, for any $S \subseteq N$, the linear supermodular inequalities [41]

$$g(S) + \sum_{i \in N \setminus S} \rho(i, S) x_i - \sum_{i \in S} \rho(i, N \setminus i)(1 - x_i) \leq t, \text{ and} \tag{3a}$$

$$g(S) + \sum_{i \in N \setminus S} \rho(i, \emptyset) x_i - \sum_{i \in S} \rho(i, S \setminus i)(1 - x_i) \leq t \tag{3b}$$

are valid for Z.

2.2 Lifted supermodular inequalities

We now describe a family of lifted supermodular inequalities, using a lifting approach similar to the ones used in [28, 47]. Let $h : \{0, 1\}^N \times \mathbb{R}^N \to \mathbb{R} \cup \{\infty\}$ be a function defined over a mixed 0-1 domain and consider its epigraph

$$H = \left\{ (x, y, t) \in \{0, 1\}^N \times \mathbb{R}^N \times \mathbb{R} : h(x, y) \leq t \right\}.$$

Observe that H allows for arbitrary constraints, which can be encoded via function h. For example, nonnegativity and complementarity constraints can be included by letting $h(x, y) = \infty$ whenever $y_i < 0$ or $y_i(1 - x_i) \neq 0$ for some $i \in N$.

For $\alpha \in \mathbb{R}^N$, define the set function $g_\alpha : \{0, 1\}^N \to \mathbb{R} \cup \{\infty, -\infty\}$ as

$$g_\alpha(x) = \min_{y \in \mathbb{R}^N} -\alpha' y + h(x, y), \tag{4}$$

and let $B \subseteq \mathbb{R}^N$ be the set of values of α for which problem (4) is bounded for all $x \in \{0, 1\}^N$, i.e.,

$$B = \left\{ \alpha \in \mathbb{R}^N : |g_\alpha(x)| < \infty, \ \forall x \in \{0, 1\}^N \right\}.$$

Although supermodularity is defined for set functions only, we propose in Definition 1 below an extension for functions involving continuous variables as well.

Definition 1 Function h is supermodular if the set function g_α defined in (4) is supermodular for all $\alpha \in B$.

Remark 1 Suppose that h does not depend on the continuous variables y, i.e., $h(x, y) = g(x)$. In this case problem (4) is unbounded unless $\alpha = 0$, i.e., $B = \{0\}$, and we find that $h(x, y)$ is supermodular if and only if $g_0(x) = g(x)$ is supermodular. Thus, Definition 1 includes the usual definition of supermodularity for set functions as a special case.   □

Proposition 2 *If function h is supermodular, then for any $\alpha \in B$ and $S \subseteq N$, the inequalities*

$$\alpha' y + g_\alpha(S) + \sum_{i \in N \setminus S} \rho_\alpha(i, S) x_i - \sum_{i \in S} \rho_\alpha(i, N \setminus i)(1 - x_i) \le t, \ and \quad (5a)$$

$$\alpha' y + g_\alpha(S) + \sum_{i \in N \setminus S} \rho_\alpha(i, \emptyset) x_i - \sum_{i \in S} \rho_\alpha(i, S \setminus i)(1 - x_i) \le t \quad (5b)$$

are valid for H, where $\rho_\alpha(i, S) = g_\alpha(S \cup i) - g_\alpha(S)$.

Proof For any $\alpha \in B$, $S \subseteq N$, and $(x, y, t) \in H$, we find

$$t - \alpha' y \ge h(x, y) - \alpha' y \ge g_\alpha(x) \ge g_\alpha(S) + \sum_{i \in N \setminus S} \rho_\alpha(i, S) x_i - \sum_{i \in S} \rho_\alpha(i, N \setminus i)(1 - x_i),$$

where the first inequality follows directly from the definition of H, the second inequality follows by minimizing $h(y) - \alpha' y$ with respect to y, and the third inequality follows from the validity of (3a). Thus, by adding $\alpha' y$ on both sides, we find that inequality (5a) is valid. The validity of (5b) is proven identically.   □

Since inequalities (5) are valid for any $\alpha \in B$, one can obtain stronger valid inequalities by optimally choosing vector α.

Theorem 1 (Lifted supermodular inequalities) *If h is supermodular, then for any $S \subseteq N$, the lifted supermodular inequalities*

$$\max_{\alpha \in B} \ g_\alpha(S) + \sum_{i \in N \setminus S} \rho_\alpha(i, S) x_i - \sum_{i \in S} \rho_\alpha(i, N \setminus i)(1 - x_i) + \alpha' y \le t, \ and \quad (6a)$$

$$\max_{\alpha \in B} \ g_\alpha(S) + \sum_{i \in N \setminus S} \rho_\alpha(i, \emptyset) x_i - \sum_{i \in S} \rho_\alpha(i, S \setminus i)(1 - x_i) + \alpha' y \le t \quad (6b)$$

are valid for H.

Observe that while inequalities (5) are linear, inequalities (6) are *nonlinear* in x and y. Moreover, each inequality (6) is convex since it is defined as a supremum of linear inequalities. In addition, if the base supermodular inequalities (3) are strong for the convex hull of epi g_α, then the lifted supermodular inequalities (6) are strong for H as well, as formalized next. Given $\alpha \in B$, define

$$G_\alpha = \left\{ (x, t) \in \{0, 1\}^N \times \mathbb{R} : g_\alpha(x) \le t \right\}.$$

Note that conv(G_α) is a polyhedron. Theorem 2 below is a direct consequence of Theorem 1 in [47].

Theorem 2 ([47]) *If inequalities (3) and bound constraints $0 \leq x \leq 1$ describe* conv(G_α) *for all $\alpha \in B$, then the lifted supermodular inequalities (6) and bound constraints $0 \leq x \leq 1$ describe cl conv(H).*

Although Definition 1 may appear to be too restrictive to arise in practice, we show in Sect. 2.3 that supermodular functions are in fact widespread in a class of well-studied problems in mixed-integer *linear* optimization. In Sect. 3 we show that several existing results for quadratic optimization with indicators can be obtained as lifted supermodular inequalities. Perhaps, more surprisingly, for the rank-one quadratic with indicators

$$h(x, y) = \begin{cases} \left(y(N^+) - y(N^-)\right)^2 & \text{if } y \geq 0 \text{ and } y_i(1 - x_i) = 0, \ \forall i \in N^+ \cup N^- \\ \infty & \text{otherwise,} \end{cases}$$

we show in Sect. 4 that conditions in Definition 1 and Theorem 2 are satisfied as well.

2.3 Supermodular inequalities and fixed-charge networks

Given $b \in \mathbb{R}$, $u \in \mathbb{R}_+^N$, and a partition $N = N^+ \cup N^- \cup A^+ \cup A^-$, define for all $x \in \{0, 1\}^N$ the *fixed-charge network set*

$$FC(x) = \left\{ y \in \mathbb{R}_+^N : y(N^+) + y(A^+) - y(A^-) - y(N^-) \leq b, \ y_i \leq u_i, \ i \in N, \right.$$
$$\left. y_i(1 - x_i) = 0, \ i \in N^+, \ y_i x_i = 0, \ i \in N^- \right\}.$$

Wolsey [54] uses $FC(x)$ to describe network structures arising in flow problems with fixed charges on the arcs: N^+ denotes the incoming arcs into a given subgraph, N^- denotes the outgoing arcs, and whereas $A^+ \cup A^-$ denotes the internal arcs in the subgraph, and b represents the supply/demand of the subgraph. Finally, define

$$h(x, y) = \begin{cases} 0 & \text{if } y \in FC(x) \\ \infty & \text{otherwise.} \end{cases}$$

Proposition 3 ([54]) *For any $\alpha \in \mathbb{R}^N$, the function*

$$v_\alpha(x) = \max_{y \in \mathbb{R}_+^N} \alpha' y - h(x, y)$$

is submodular.

It follows that the function $g_\alpha(x) = -v_\alpha(x) = \min_{y \in \mathbb{R}_+^N} -\alpha' y + h(x, y)$ is supermodular, and inequalities (5) and (6) are valid. Moreover, Wolsey [54] shows that the linear supermodular inequalities (5) with $\alpha \in \{-1, 0, 1\}^N$ include as special

cases well-known inequalities for mixed-integer linear optimization such as flow-cover inequalities [45, 50] and inequalities for capacitated lot-sizing [9, 46]; several other classes for fixed-charge network flow problems are special cases as well [4, 11, 12]. Therefore, the inequalities presented in this paper can be interpreted as nonlinear generalizations of the aforementioned inequalities.

3 Previous results as lifted supermodular inequalities

In order to illustrate the approach, in this section, we show how existing results for quadratic optimization with indicators can be derived using the lifted supermodular inequalities (6).

3.1 The single-variable case

Consider, first, the single-variable case

$$X^1 = \left\{(x, y, t) \in \{0, 1\} \times \mathbb{R}_+ \times \mathbb{R} : y^2 \leq t, \ y(1 - x) = 0\right\}$$

for which cl conv(X^1) is given by the perspective reformulation [2, 19, 23, 32]:

$$\text{cl conv}(X^1) = \left\{(x, y, t) \in [0, 1] \times \mathbb{R}_+ \times (\mathbb{R} \cup \infty) : \frac{y^2}{x} \leq t\right\}.$$

Note that cl conv$(X^1) \subseteq \mathbb{R}^2 \times (\mathbb{R} \cup \infty)$. We now derive the perspective reformulation as a special case, in fact, using a modular inequality. Note that $g_\alpha(0) = 0$ and $g_\alpha(1) = \min_{y \in \mathbb{R}_+} -\alpha y + y^2 = -\frac{\alpha_+^2}{4}$ since $y^* = \alpha/2$ if $\alpha \geq 0$ and $y^* = 0$ otherwise. Thus, g_α is a modular function for any $\alpha \in \mathbb{R}^N$, and inequalities (3) reduce to

$$t \geq -\frac{1}{4}\alpha_+^2 x.$$

Then, we find that inequalities (6) reduce to the perspective of y^2:

$$t \geq \max_{\alpha \in \mathbb{R}^N} -\frac{1}{4}\alpha_+^2 x + \alpha y = \frac{y^2}{x}. \qquad\qquad \text{(with } \alpha^* = 2y/x\text{)}$$

3.2 The rank-one case with free continuous variables

Consider the relaxation of X obtained by dropping the non-negativity constraints $y \geq 0$:

$$X_f = \left\{ (x, y, t) \in \{0, 1\}^N \times \mathbb{R}^N \times \mathbb{R} : y(N)^2 \leq t, \ y_i(1 - x_i) = 0, \ \forall i \in N \right\}.$$

Observe that any rank-one quadratic constraint of the form $\left(\sum_{i \in N} c_i y_i \right)^2 \leq t$ with $c_i \neq 0$ can be transformed into the form given in X_f by scaling the continuous variables (so that $|c_i| = 1$) and negating variables as $\bar{y}_i := -y_i$ if $c_i < 0$. The closure of the convex hull of X_f is derived in [7], and the effectiveness of the resulting inequalities is demonstrated on sparse regression problems. We now re-derive the description of cl conv(X_f) using lifted supermodular inequalities.

For $S \subseteq N$, we have

$$g_\alpha(S) = \min_{y \in \mathbb{R}^S} -\alpha' y + y(S)^2.$$

It is easy to see that $g_\alpha(x_S) = -\infty$ unless $\alpha_i = \alpha_j$ for all $i \neq j$, see [7]. Therefore, letting $\bar{\alpha} = \alpha_i$ for all $i \in N$, we find that

$$g_{\bar{\alpha}}(x_S) = \min_{y \in \mathbb{R}^S} -\bar{\alpha} y(S) + y(S)^2 = \begin{cases} 0 & \text{if } S = \emptyset \\ -\bar{\alpha}^2/4 & \text{otherwise,} \end{cases}$$

where the optimal solution is found by setting $y(S) = \bar{\alpha}/2$. The function g_α is supermodular since $\rho_{\bar{\alpha}}(i, \emptyset) = -\bar{\alpha}^2/4$ and $\rho_{\bar{\alpha}}(i, S) = 0$ for any $S \neq \emptyset$.

Letting $S = \{1\}$, inequality (6a) reduces to

$$\max_{\bar{\alpha} \in \mathbb{R}} -\frac{\bar{\alpha}^2}{4} + \bar{\alpha} y(N) \leq t \ \Leftrightarrow \ y(N)^2 \leq t. \qquad \text{(with } \bar{\alpha} = 2y(N))$$

Also letting $S = \{1\}$, inequality (6b) reduces to

$$\max_{\bar{\alpha} \in \mathbb{R}} -\frac{\bar{\alpha}^2}{4} x(N) + \bar{\alpha} y(N) \leq t \ \Leftrightarrow \ \frac{y(N)^2}{x(N)} \leq t. \qquad \text{(with } \bar{\alpha} = 2y(N)/x(N))$$

These two supermodular inequalities are indeed sufficient to describe conv(X_f) [7]. As we shall see in Sect. 4, incorporating the non-negativity constraints $y \geq 0$, conv(X) is substantially more complex than conv(X_f). Nonetheless, as shown in Example 1, the resulting convexification is substantially stronger as well.

3.3 The rank-one case with a negative off-diagonal

Consider the special case of X with two continuous variables ($N = \{1, 2\}$) with a negative off-diagonal:

$$X_-^2 = \left\{(x, y, t) \in \{0, 1\}^2 \times \mathbb{R}_+^2 \times \mathbb{R} : (y_1 - y_2)^2 \leq t, \ y_i(1 - x_i) = 0, \ i = 1, 2\right\}.$$

Observe that any quadratic constraint of the form $(c_1 y_1 - c_2 y_2)^2 \leq t$ with $c_1, c_2 > 0$ can be written as in X_-^2 by scaling the continuous variables.

For $\alpha \in \mathbb{R}^2$, observe that if $\alpha_1 + \alpha_2 > 0$,

$$g_\alpha(\{1, 2\}) = \min_{y \in \mathbb{R}_+^2} -\alpha_1 y_1 - \alpha_2 y_2 + (y_1 - y_2)^2$$

is unbounded. Otherwise,

$$g_\alpha(\emptyset) = 0,$$

$$g_\alpha(\{1\}) = -\frac{\alpha_1^2}{4} \text{ if } \alpha_1 \geq 0 \text{ and } g_\alpha(\{1\}) = 0 \text{ otherwise,}$$

$$g_\alpha(\{2\}) = -\frac{\alpha_2^2}{4} \text{ if } \alpha_2 \geq 0 \text{ and } g_\alpha(\{2\}) = 0 \text{ otherwise,}$$

$$g_\alpha(\{1, 2\}) = \begin{cases} -\frac{\alpha_1^2}{4} & \text{if } \alpha_1 \geq 0 \\ -\frac{\alpha_2^2}{4} & \text{if } \alpha_2 \geq 0 \\ 0 & \text{if } \alpha_1 \leq 0 \text{ and } \alpha_2 \leq 0. \end{cases}$$

In particular, g_α is supermodular (and in fact modular) for any fixed α such that $\alpha_1 + \alpha_2 \leq 0$: for any $i = 1, 2$ and $S \subseteq N \setminus i$, $\rho_\alpha(i, S) = -\frac{\max\{0, \alpha_i\}^2}{4}$. Letting $S = \emptyset$, inequality (6a) reduces to

$$\max_{\alpha_1 + \alpha_2 \leq 0} -\frac{\max\{0, \alpha_1\}^2}{4} x_1 - \frac{\max\{0, \alpha_2\}^2}{4} x_2 + \alpha_1 y_1 + \alpha_2 y_2 \leq t. \tag{7}$$

An optimal solution of (7) can be found as follows. If $y_1 \geq y_2$, then set $\alpha_1 > 0$ and $\alpha_2 = -\alpha_1 < 0$. Moreover, in this case, the optimal value is given by

$$\max -\frac{\alpha_1^2}{4} x_1 + \alpha_1(y_1 - y_2) = \frac{(y_1 - y_2)^2}{x_1}.$$

The case $y_2 \geq y_1$ is identical. The resulting piecewise valid inequality

$$t \geq \begin{cases} \frac{(y_1 - y_2)^2}{x_1} & \text{if } y_1 \geq y_2 \\ \frac{(y_1 - y_2)^2}{x_2} & \text{if } y_2 \geq y_1 \end{cases} \tag{8}$$

along with the bound constraints $0 \leq x \leq 1$, $0 \leq y$ describe cl conv(X_-^2) [6]. We point that a conic quadratic representation for cl conv(X_-^2) and generalizations to (not necessarily rank-one) quadratic functions with negative off-diagonals are given in [13].

3.4 Outlier detection with temporal data

In the context of outlier detection with temporal data, Gómez [30] studies the set

$$X_T = \left\{ (x, y, t) \in \{0, 1\}^2 \times \mathbb{R}^4 \times \mathbb{R} : \frac{a_1}{2}(y_3 - y_1)^2 + (y_3 - y_4)^2 + \frac{a_2}{2}(y_4 - y_2)^2 \leq t, \right.$$
$$\left. y_1(1 - x_1) = 0, \ y_2(1 - x_2) = 0 \right\}$$

where $a_1, a_2 > 0$ are constants. While we refer the reader to [30] for details on the derivation of cl conv(X_T), we point out that it can in fact be described by lifted supermodular inequalities. Indeed, in this case, function g_α is given by

$$g_\alpha(x) = K_1(\alpha) - K_2(\alpha) \max\{x_1, x_2\},$$

where $K_1(\alpha)$ and $K_2(\alpha)$ are constants that do not depend on x and $K_2(\alpha) \geq 0$. Since $\max\{x_1, x_2\}$ is a submodular function, it follows that g_α is supermodular.

4 Convex hull via lifted supermodular inequalities

We now turn our attention to the rank-one sets X and X_+. This section is devoted to showing that the lifted supermodular inequalities (6) are sufficient to describe cl conv(X) and cl conv(X_+). By Theorem 2, it suffices to derive an explicit form of the projection function g_α and show that inequalities (3) describe the convex hull of its epigraph G_α. The rest of this section is organized as follows. In Sect. 4.1 we derive the set function g_α defined in (4) for the rank-one quadratic function and then show that it is supermodular. In Sect. 4.2 we describe the convex hull of G_α using only a small subset of the supermodular inequalities (3).

4.1 The set function g_α

We present the derivation of set function g_α for X_+ and X separately, and then verify that g_α is indeed supermodular.

4.1.1 Derivation for X_+

For X_+,

$$h(x, y) = \begin{cases} y(N)^2 & \text{if } y \geq 0 \text{ and } y_i(1 - x_i) = 0, \ \forall i \in N \\ \infty & \text{otherwise.} \end{cases}$$

Therefore, for $S \subseteq N$,

$$g_\alpha(x_S) = \min_{y \in \mathbb{R}_+^S} -\alpha_S' y_S + y(S)^2. \tag{9}$$

Note that (9) is bounded for all $\alpha \in \mathbb{R}^S$, thus $B = \mathbb{R}^N$. Since, for $\alpha_i < 0$, $y_i = 0$ in any optimal solution, we assume for simplicity that $\alpha \geq 0$ and $B = \mathbb{R}_+^N$. From the KKT conditions corresponding to variable $y_k \geq 0$ in (9), we find that

$$2y(S) \geq \alpha_k, \tag{10}$$

and, by complementary slackness, (10) holds at equality whenever $y_k > 0$. Moreover, let $j \in S$ such that $\alpha_j = \max_\alpha(S)$; setting $y_j = \alpha_j/2$ and $y_i = 0$ for $i \in S \backslash j$, we find a feasible solution for (9) that satisfies all dual feasibility conditions (10) and complementary slackness, and therefore is optimal for the convex optimization problem (9). Thus, we conclude that

$$g_\alpha(x_S) = -\frac{\max_\alpha(S)^2}{4}.$$

4.1.2 Derivation for X

For the general case of X,

$$h(x, y) = \begin{cases} \left(y(N^+) - y(N^-)\right)^2 & \text{if } y \geq 0 \text{ and } y_i(1 - x_i) = 0, \ \forall i \in N^+ \cup N^- \\ \infty & \text{otherwise.} \end{cases}$$

Therefore, for $S \subseteq N^+ \cup N^-$,

$$g_\alpha(x_S) = \min_{y \in \mathbb{R}_+^S} -\alpha' y + \left(y(N^+ \cap S) - y(N^- \cap S)\right)^2. \tag{11}$$

If $S \cap N^- = \emptyset$ or $S \cap N^+ = \emptyset$, then we find from Sect. 4.1.1 that $g_\alpha(x_S) = -\max_\alpha(S)^2/4$. Now let $S^+ := S \cap N^+$ and $S^- := S \cap N^-$, and assume $S^+ \neq \emptyset$ and $S^- \neq \emptyset$. We first state conditions under which (11) is bounded, and then we provide the explicit description of g_α.

Lemma 1 *Problem* (11) *is bounded if and only if*

$$\max_\alpha(S^+) \leq -\max_\alpha(S^-). \tag{12}$$

Proof Let $p = \arg\max_{i \in S^+} \alpha_i$ and $q = \arg\max_{i \in S^-} \alpha_i$. If $\alpha_p + \alpha_q > 0$, then $e_p + e_q$ is an unbounded direction. Otherwise,

$$-\alpha' y + \left(y(S^+) - y(S^-)\right)^2 \geq -\alpha_p y(S^+) - \alpha_q y(S^-) + \left(y(S^+) - y(S^-)\right)^2$$

$$\geq -\alpha_p(y(S^+) - y(S^-)) + \left(y(S^+) - y(S^-)\right)^2$$
$$\geq -\alpha_p^2/4,$$

where the second inequality follows from $\alpha_p + \alpha_q \leq 0$. $\qquad\square$

Note that we may equivalently rewrite (12) as $\alpha_i + \alpha_j \leq 0$, for all $i \in S^+$, $j \in S^-$, and in particular,

$$B = \left\{\alpha \in \mathbb{R}^N : \alpha_i + \alpha_j \leq 0 \text{ for all } i \in N^+, j \in N^-\right\}.$$

Proposition 4 *Function g_α is given by*

$$g_\alpha(x_S) = \begin{cases} 0 & \text{if } \alpha \leq 0 \\ -\max_\alpha(S^+)^2/4 & \text{if } \alpha \nleq 0, (12) \text{ and } \alpha_i \leq 0 \text{ for all } i \in S^- \\ -\max_\alpha(S^-)^2/4 & \text{if } \alpha \nleq 0, (12) \text{ and } \alpha_i \leq 0 \text{ for all } i \in S^+ \\ -\infty & \text{otherwise.} \end{cases}$$

Proof If $\alpha \leq 0$, then $g_\alpha(x_S) \geq 0$ and the lower bound can be obtained by setting $y = 0$. We now assume $y \nleq 0$. Note that for (12) to hold, if there exists $j \in S^-$ such that $\alpha_j \geq 0$, then $\alpha_i \leq 0$ for all $i \in S^+$, and vice versa. Therefore, either $\alpha_i \leq 0$ for all $i \in S^+$ or $\alpha_j \leq 0$ for all $j \in S^-$.

First, assume that $\alpha_j \leq 0$ for all $j \in S^-$. In this case, there exists an optimal solution of (11) where $y(S^-) = 0$ and (11) reduces to (9). Then, we may assume that $\alpha_i \geq 0$ for all $i \in S^+$ as in Sect. 4.1.1, and arrive at

$$g_\alpha(x_S) = -\frac{\max_\alpha(S^+)^2}{4}.$$

By symmetry, if $\alpha_i \leq 0$ for all $i \in S^+$, we may assume that $\alpha_j \geq 0$ for all $i \in S^-$ and

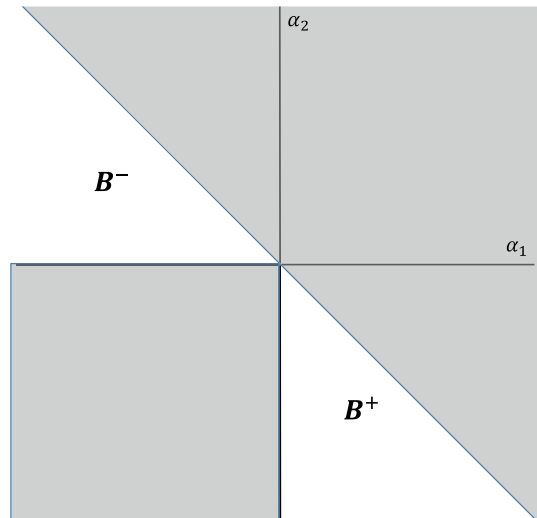$$g_\alpha(x_S) = -\frac{\max_\alpha(S^-)^2}{4}.$$

$\qquad\square$

Observe that if $\alpha_i \leq 0$ for all $i \in S^-$ and there exists $\alpha_j \in S^+$ such that $\alpha_j < 0$, then setting $\alpha_j = 0$ does not change the function g_α. Thus we can assume without loss of generality in optimization problem (6) that

$$B = \left\{\alpha \in \mathbb{R}^N : \alpha_i\alpha_j \leq 0 \text{ and } \alpha_i + \alpha_j \leq 0 \text{ for all } i \in N^+, j \in N^-\right\}.$$

It is convenient to partition B into two sets so that $B = B^+ \cup B^-$, where

$$B^+ = \left\{\alpha \in \mathbb{R}^N : \alpha_i \geq 0 \,\forall i \in N^+, \alpha_j \leq 0 \,\forall j \in N^-, \text{ and } \alpha_i + \alpha_j \leq 0 \,\forall i \in N^+, j \in N^-\right\}$$

Fig. 2 Depiction of B^+ and B^- in a two-dimensional example with $N^+ = \{1\}$ and $N^- = \{2\}$. The upper right shaded region (triangle) corresponds to the region where $g_\alpha(x) = -\infty$; the lower left shaded region (square) corresponds to the region discarded, as equivalent solutions of (6) can be found in either B^+ or B^-

$$B^- = \left\{ \alpha \in \mathbb{R}^N : \alpha_i \leq 0 \; \forall i \in N^+, \alpha_j \geq 0 \; \forall j \in N^-, \text{ and } \alpha_i + \alpha_j \leq 0 \; \forall i \in N^+, j \in N^- \right\}$$

and analyze the inequalities separately for each set. Figure 2 depicts regions B^+ and B^- for a two-dimensional case.

Therefore, instead of studying inequalities (6) directly, one can equivalently study their relaxation where either $\alpha \in B^+$ or $\alpha \in B^-$; consequently, each inequality (6) corresponds to (the maximum of) two simpler inequalities. Since the sets B^+ and B^- are symmetric, and inequalities (6) corresponding to $\alpha \in B^-$ are simply inequalities where the role of N^+ and N^- is interchanged (and $\alpha \in B^+$), the analysis and derivation of the inequalities is simplified. Therefore, in the sequel, we will derive the inequalities for $\alpha \in B^+$ only and then state the inequalities corresponding to B^- by interchanging N^+ and N^-.

Supermodularity

For $\alpha \in B^+$, the set function $g_\alpha(x)$ for X is monotone non-increasing, also it is supermodular as $\max_\alpha(S^+)$ is submodular. The case for $\alpha \in B^-$ is analogous.

4.2 Convex hull of epi g_α

In this section we show that a small subset of the supermodular inequalities (3a) are sufficient to describe the convex hull of the epigraph of the set function g_α, i.e.,

$$G_\alpha = \left\{ (x, t) \in \{0, 1\}^N \times \mathbb{R} : -\frac{\max_{i \in N}\{\alpha_i^2 x_i\}}{4} \leq t \right\}, \tag{13}$$

where $\alpha \geq 0$ – observe that since x is binary, $(\max_{i \in N}\{\alpha_i x_i\})^2 = \max_{i \in N}\{\alpha_i^2 x_i\}$.

Given nonempty $S \subseteq N$, $\ell \in \arg\max_{i \in S}\{\alpha_i\}$, $k \in \arg\max_{i \in N \setminus \ell}\{\alpha_i\}$, and $T = \{i \in N \setminus S : \alpha_i > \alpha_\ell\}$; observe that $T = \emptyset$ if and only if $\alpha_\ell \geq \alpha_k$. Then, valid inequalities (3a) for G_α reduce to

$$
t \geq \begin{cases} -\frac{\alpha_\ell^2}{4} + \frac{\alpha_\ell^2 - \alpha_k^2}{4}(1 - x_\ell) & \text{if} \quad \alpha_\ell \geq \alpha_k \\ -\frac{\alpha_\ell^2}{4} - \sum\limits_{i \in T} \frac{\alpha_i^2 - \alpha_\ell^2}{4} x_i & \text{if} \quad \alpha_\ell \leq \alpha_k. \end{cases} \tag{14}
$$

If $S = \emptyset$, then valid inequalities (3a) reduce to

$$
t \geq -\sum_{i \in N} \frac{\alpha_i^2}{4} x_i.
$$

Remark 2 Observe that if $\alpha_\ell \geq \alpha_k$, then the inequality

$$
t \geq -\frac{\alpha_\ell^2}{4} + \frac{\alpha_\ell^2 - \alpha_k^2}{4}(1 - x_\ell) = -\frac{\alpha_k^2}{4} - \frac{\alpha_\ell^2 - \alpha_k^2}{4} x_\ell
$$

can also be obtained by setting $S = N \setminus \ell$ (or by choosing any $S \subseteq N \setminus \ell$ such that $k \in S$). Therefore, when considering inequalities (14), we can assume without loss of generality that there exists $k \in \arg\max_{i \in N}\{\alpha_i\}$ such that $k \notin S$ and, thus, the case $\alpha_\ell \geq \alpha_k$ can be ignored. □

Remark 3 Suppose that the variables are indexed such that $\alpha_1 \leq \cdots \leq \alpha_n$, let $\alpha_0 = 0$, and let $\ell = \max_{i \in S}\{i\}$ if $S \neq \emptyset$ and $\ell = 0$ otherwise. Observe that we can assume without loss of generality that $i \in S$ for all $i \leq \ell$, since inequalities (14) are the same whether $i \in S$ or not. Therefore, it follows that there are only n inequalities (14) given by

$$
t \geq -\frac{\alpha_\ell^2}{4} - \sum_{i=\ell+1}^{n} \frac{\alpha_i^2 - \alpha_\ell^2}{4} x_i, \quad \ell = 0, \ldots, n-1. \tag{15}
$$

□

We now show that inequalities (14) characterize the convex hull of G_α.

Proposition 5 *Inequalities* (14) *and bound constraints describe* $conv(G_\alpha)$.

Proof Let $(x, t) \in [0, 1]^N \times \mathbb{R}$. By definition, $(x, t) \in conv(G_\alpha)$ if and only

$$
t \geq \min_{\lambda} - \sum_{S \subseteq N} \frac{\max_\alpha(S)^2}{4} \lambda_S \tag{16a}
$$

$$
\text{s.t.} \sum_{S \subseteq N : i \in S} \lambda_S = x_i, \quad i \in N \tag{16b}
$$

$$\sum_{S \subseteq N} \lambda_S = 1 \tag{16c}$$

$$\lambda_S \geq 0, \quad S \subseteq N, \tag{16d}$$

where constraints (16b) can be restated as $x = \sum_{i \in S} \lambda_S x_S$. From linear programming duality, we find the equivalent condition

$$t \geq \max_{\mu, \gamma} \sum_{i \in N} x_i \mu_i + \gamma \tag{17a}$$

$$\text{s.t.} \sum_{i \in S} \mu_i + \gamma \leq -\frac{\max_\alpha(S)^2}{4}, \quad S \subseteq N \tag{17b}$$

$$\mu \in \mathbb{R}^N, \ \gamma \in \mathbb{R}. \tag{17c}$$

Any feasible solution (μ, γ) of (17) yields a valid inequality for $\text{conv}(G_\alpha)$. Moreover, characterizing the optimal solutions of (17) (for all $x \in [0, 1]^N$) results in the convex hull description of G_α.

Suppose, without loss of generality, that $\alpha_1 \leq \ldots \leq \alpha_n$, let $\alpha_0 = 0$, and let $\ell \in \{0, \ldots, n-1\}$ be the smallest index such that $\sum_{i=\ell+1}^n x_i \leq 1$; thus, if $\ell > 0$, then $\sum_{i=\ell}^n x_i > 1$. We claim that the dual solution given by $\hat{\gamma} = -\frac{\alpha_\ell^2}{4}$, $\hat{\mu}_i = 0$ for $i \leq \ell$ and $\hat{\mu}_i = -\frac{\alpha_i^2 - \alpha_\ell^2}{4}$ for $i > \ell$ is optimal for (17).

First, we verify that $(\hat{\mu}, \hat{\gamma})$ is feasible for (17). Observe that for any $S \subseteq \{1, \ldots, \ell\}$, constraint (17b) reduces to $-\frac{\alpha_\ell^2}{4} \leq -\frac{\max_\alpha(S)^2}{4}$, which is indeed satisfied. For any S such that the maximum element $j > \ell$, we find that (17b) reduces to $\sum_{i \in S: i \neq j} \hat{\mu}_i \leq 0$; since $\hat{\mu} \leq 0$, the constraint is satisfied. For $S = \emptyset$, constraint (17b) reduces to $\gamma \leq 0$, which is satisfied. To verify complementary slackness (later), note that constraints (17b) corresponding to sets (a) $S = T \cup \{j\}$, where $T \subseteq \{1, \ldots, \ell\}$ and $j > \ell$ (i.e., containing exactly one element greater than ℓ), and (b) $S = T \cup \{\ell\}$, where $T \subseteq \{1, \ldots, \ell-1\}$ (i.e., containing ℓ but no greater element) are satisfied at equality.

Finally, for $(\hat{\mu}, \hat{\gamma})$, the objective function (17a) is of the form (15):

$$t \geq -\frac{\alpha_\ell^2}{4} - \sum_{i=\ell+1}^n \frac{\alpha_i^2 - \alpha_\ell^2}{4} x_i.$$

To verify that $(\hat{\mu}, \hat{\gamma})$ is optimal for (17), we construct a primal solution $\hat{\lambda}$ feasible for (16) satisfying complementary slackness. The greedy algorithm for constructing $\hat{\lambda}$ is presented in Algorithm 1 and illustrated with an example in Fig. 3.

We now check that constraint (16c) is satisfied. At the end of the algorithm, $\sum_{S \subseteq N} \hat{\lambda}_S = \Lambda$ (since variable Λ is updated each time $\hat{\lambda}$ is updated). Moreover, at the end of the first cycle (line 13) we have $\Lambda = \sum_{i=\ell+1}^n x_i$. If $\ell = 0$, then $\Lambda = 1$ trivially (line 16); otherwise, at the end of the second cycle (line 22) and additional

(A) Buffer $1 - x_4 - x_5 = 0.1$
for index $\ell = 3$, $\hat{\lambda} = 0$

(B) $\hat{\lambda}_{1,2,3,5} \leftarrow x_2 = 0.2$. Constraint (16b) for x_2 is met

(C) $\hat{\lambda}_{1,3,5} \leftarrow x_5 - x_2 = 0.1$. Constraint (16b) for x_5 is met

(D) $\hat{\lambda}_{1,3,4} \leftarrow 0.1$. Constraint (16b) for x_3 is not met due to the buffer

(E) $\hat{\lambda}_{1,4} \leftarrow 0.5$. Constraint (16b) for x_4 is met and the buffer is removed

(F) $\hat{\lambda}_{1,3} \leftarrow 0.1$. Constraints (16b) for x_1, x_3 and (16c) are met
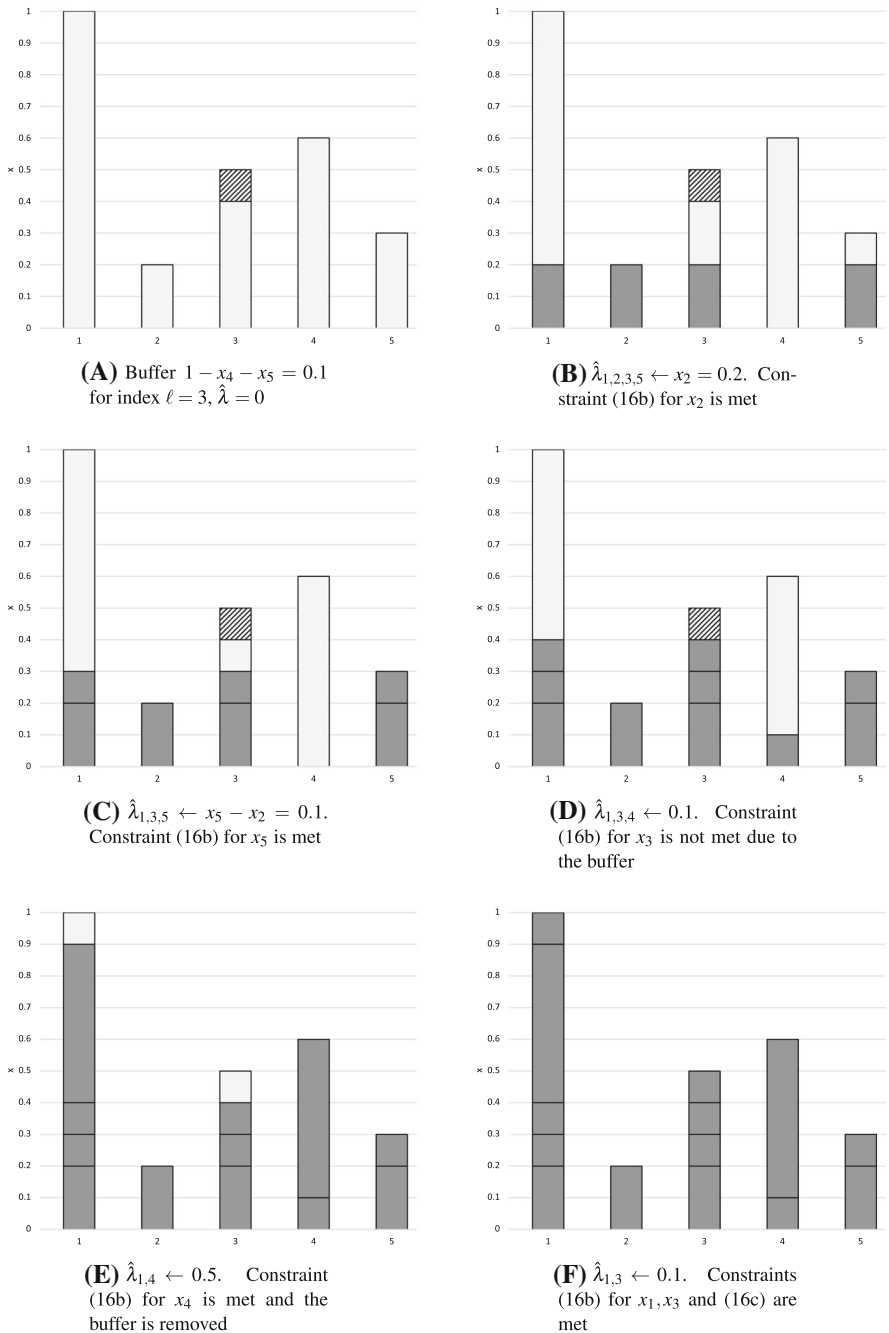
Fig. 3 Algorithm 1 with $x = (1, 0.2, 0.5, 0.6, 0.3)$ and $\ell = 3$

Algorithm 1 Algorithm for problem (16)

Input: $x_1, \ldots, x_n$ with $0 =: \alpha_0 \leq \alpha_1 \leq \cdots \leq \alpha_n$.
Output: $\hat{\lambda}$ optimal for (16).
1: $\ell \leftarrow \min\{k : \sum_{i=k+1}^{n} x_i \leq 1\}$
2: $\hat{\lambda}_S \leftarrow 0$ for all $S \subseteq N$
3: $\hat{x}_i \leftarrow x_i$ for all $i \in N \backslash \ell$
4: **if** $\ell > 0$ **then**
5:     $\hat{x}_\ell \leftarrow x_\ell - \left(1 - \sum_{i=\ell+1}^{n} x_i\right)$ ▷ Buffer for ℓ; $\hat{x}_\ell > 0$ by definition of ℓ
6: **end if**
7: $\Lambda \leftarrow 0$ ▷ Variable for $\sum_S \hat{\lambda}_S$
8: **for** $j = n \ldots \ell + 1$ **do**
9:     **while** $\hat{x}_j > 0$ **do**
10:         $S \leftarrow \{i \leq \ell : \hat{x}_i > 0\} \cup j$
11:         ALLOCATE(S)
12:     **end while**
13: **end for**   ▷ At this point $\Lambda = \sum_{i=\ell+1}^{n} x_i$
14: **if** $\ell = 0$ **then**
15:     $\hat{\lambda}_\emptyset \leftarrow 1 - \Lambda$ ▷ Constraint (16c) is satisfied
16:     $\Lambda \leftarrow 1$
17: **else**
18:     $\hat{x}_\ell \leftarrow 1 - \sum_{i=\ell+1}^{n} x_i$ ▷ Buffer is removed
19:     **while** $\hat{x}_\ell > 0$ **do**
20:         $S \leftarrow \{i < \ell : \hat{x}_i > 0\} \cup \ell$
21:         ALLOCATE(S)
22:     **end while**   ▷ $\Lambda = 1$
23: **end if**
24: **return** $\hat{\lambda}$
25:
26: **function** ALLOCATE(S)
27:     $v \leftarrow \min_{i \in S} \hat{x}_i$
28:     $\Lambda \leftarrow \Lambda + v$
29:     $\hat{\lambda}_S \leftarrow v$
30:     $\hat{x}_i \leftarrow \hat{x}_i - v$ for all $i \in S$ ▷ At this point $\hat{x}_i = (x_i - \Lambda)_+$ for all $i < \ell$
31: **end function**

---

value of $\hat{x}_\ell = 1 - \sum_{i=\ell+1}^{n} x_i$ (line 18) is added to Λ. Hence, at the end of the algorithm

$$\Lambda = \sum_{S \subseteq N} \hat{\lambda}_S = \sum_{i=\ell+1}^{n} x_i + \left(1 - \sum_{i=\ell+1}^{n} x_i\right) = 1.$$

Next, we verify that constraints (16b) are satisfied. For $i \in \{1, \ldots, \ell - 1\}$, at any point in the algorithm, we have that $\sum_{S \subseteq N: i \in S} \lambda_S = x_i - \hat{x}_i$. Since, at any point, $\hat{x}_i = (x_i - \Lambda)_+$ and $\Lambda = 1$ at the end of the algorithm, it follows that $\sum_{S \subseteq N: i \in S} \lambda_S = x_i$. For $i \in \{\ell + 1, \ldots, n\}$ we also have that $\sum_{S \subseteq N: i \in S} \lambda_S = x_i - \hat{x}_i$, and $\hat{x}_i = 0$ at the end (line 13). Finally, for $i = \ell > 0$, we have that

$$\sum_{S \subseteq N: \ell \in S} \lambda_S = \left(x_\ell - \left(1 - \sum_{i=\ell+1}^{n} x_i\right)\right) + \left(1 - \sum_{i=\ell+1}^{n} x_i\right) = x_\ell.$$

Finally, to check that $\hat{\lambda}$ satisfies complementary slackness, it suffices to observe that all updates of $\hat{\lambda}$ correspond to sets S such that exactly one element of S is greater than ℓ (line 10), or to sets S with no element greater than ℓ and where $\ell \in S$ (line 20), where the corresponding dual constraints are satisfied at equality.

Therefore, we conclude that $\hat{\lambda}$ and $(\hat{\mu}, \hat{\gamma})$ are an optimal primal-dual pair. Since problem (17) admits for any $x \in [0, 1]$ an optimal solution of the form (15), it follows that those inequalities and bound constraints describe conv(G_α). □

Finally, we obtain the main result of this section: that the (nonlinear) lifted super-modular inequalities

$$t \geq \max_{\alpha \in B^+} -\frac{\max_\alpha(S^+)^2}{4} - \sum_{i \in N^+ \setminus S^+} \frac{(\alpha_i^2 - \max_\alpha(S^+)^2)_+}{4} x_i + \alpha' y, \quad \forall S^+ \subseteq N^+ \quad (18)$$

$$t \geq \max_{\alpha \in B^-} -\frac{\max_\alpha(S^-)^2}{4} - \sum_{i \in N^- \setminus S^-} \frac{(\alpha_i^2 - \max_\alpha(S^-)^2)_+}{4} x_i + \alpha' y, \quad \forall S^- \subseteq N^- \quad (19)$$

are sufficient to describe the closure of the convex hull of X.

Proposition 6 *Lifted supermodular inequalities* (18)–(19) *and the bound constraints* $0 \leq x \leq 1$, $y \geq 0$ *describe cl conv(X).*

Proof Follows immediately from Proposition 5 and Theorem 2. □

Remark 4 We end this section with the remark that optimization of a linear function over X can be done easily using the projection function g_α. Consider

$$\min \left\{ -\alpha' y + \beta' x + t : (x, y, t) \in X \right\}.$$

Projecting out the continuous variables using g_α, the problem reduces to

$$\min_{x \in \{0,1\}^N} \beta' x - \max_{i \in N} \{\alpha_i^2 x_i\}/4. \quad (20)$$

Assume without loss of generality that $\beta \geq 0$ (otherwise, set $x_i = 1$ whenever $\beta_i < 0$). Then an optimal solution of (20) corresponds to either setting $x = 0$, or setting a single variable $x_i = 1$ where $i \in \arg\max_{i \in N} \beta_i + \alpha_i/4$. Identifying such an index can be done in $O(n)$. □

5 Explicit form of the lifted supermodular inequalities

In this section we derive explicit forms of the lifted supermodular inequalities (18)–(19). In Sect. 5.1 we describe the inequalities in the original space of variables, and describe how to solve the separation problem. In Sect. 5.2 we provide conic quadratic representable inequalities in an extended space, which can then be implemented with off-the-shelf conic solvers.

5.1 Inequalities and separation in the original space of variables

5.1.1 Lifted inequalities for *X*

We first present the inequalities for the more general set X. Finding a closed form expression for the lifted supermodular inequalities (18) for all $S^+ \subseteq N^+$ amounts to solving the maximum lifting problem

$$t \geq \max_{S^+ \subseteq N^+, \alpha \in B^+} -\frac{\max_\alpha (S^+)^2}{4} - \sum_{i \in N^+ \setminus S^+} \frac{\left(\alpha_i^2 - \max_\alpha (S^+)^2\right)_+}{4} x_i + \alpha' y. \quad (21)$$

We now give a closed form expression for (21). Let $m = |N^+|$, and given $(\bar{x}, \bar{y}) \in [0, 1]^N \times \mathbb{R}_+^N$, index variables in N^+ so that $\bar{y}_{(1)}/\bar{x}_{(1)} \leq \bar{y}_{(2)}/\bar{x}_{(2)} \leq \cdots \leq \bar{y}_{(m)}/\bar{x}_{(m)}$.

Proposition 7 *Given* $(\bar{x}, \bar{y}, \bar{t}) \in [0, 1]^N \times \mathbb{R}_+^N \times \mathbb{R}$, *if there exist indexes* $0 \leq \kappa_1 < \kappa_2 \leq m + 1$ *such that the (possibly empty) sets* $L = \left\{(i) \in N^+ : i \leq \kappa_1\right\}$ *and* $U = \left\{(i) \in N^+ : i \geq \kappa_2\right\}$ *satisfy*

$$1 - \bar{x}(N^+ \setminus L) \geq 0 \tag{22a}$$

$$\frac{\bar{y}(L)}{1 - \bar{x}(N^+ \setminus L)} < \frac{\bar{y}_i}{\bar{x}_i}, \quad i \in N^+ \setminus L \tag{22b}$$

$$\frac{\bar{y}(L)}{1 - \bar{x}(N^+ \setminus L)} \geq \frac{\bar{y}_i}{\bar{x}_i}, \quad i \in L \tag{22c}$$

$$\bar{y}(U) - \bar{y}(N^-) \geq 0 \tag{22d}$$

$$\frac{\bar{y}(U) - \bar{y}(N^-)}{\bar{x}(U)} > \frac{y_i}{x_i}, \quad i \in N^+ \setminus U \tag{22e}$$

$$\frac{\bar{y}(U) - \bar{y}(N^-)}{\bar{x}(U)} \leq \frac{\bar{y}_i}{\bar{x}_i}, \quad i \in U \tag{22f}$$

$$\frac{\bar{y}(L)}{1 - \bar{x}(N^+ \setminus L)} < \frac{\bar{y}(U) - \bar{y}(N^-)}{\bar{x}(U)}, \tag{22g}$$

then inequality (21) *is satisfied if and only if*

$$\bar{t} \geq \frac{\bar{y}(L)^2}{1 - \bar{x}(N^+ \setminus L)} + \sum_{i \in N^+ \setminus (L \cup U)} \frac{\bar{y}_i^2}{\bar{x}_i} + \frac{\left(\bar{y}(U) - \bar{y}(N^-)\right)^2}{\bar{x}(U)}; \tag{23}$$

otherwise, inequality (21) *is satisfied if and only if* $\bar{t} \geq \left(\bar{y}(N^+) - \bar{y}(N^-)\right)^2$.

Below we state two remarks on Proposition 7, and then we prove the result.

Remark 5 Inequalities (23), when sets L and U are fixed, are neither valid for cl conv(X) nor convex for all $(x, y) \in [0, 1]^N \times \mathbb{R}_+^N$. Indeed, if condition (22a) is not satisfied, then (23) may not be convex. Moreover, suppose that $L = \{j\}$ and

$U = \{k\}$ for some $j, k \in S^+$: note that setting $x_i = y_i = 0$ for all $i \in N \setminus \{j, k\}$, $x_j = x_k = 1$, $y_j, y_k > 0$, and $t = (y_j + y_k)^2$ is feasible for X, but this point is cut off by inequality (23) since $\frac{y(L)^2}{1-x(N^+\setminus L)} = \frac{y_j^2}{1-x_k} = \infty$.

In fact, if $(x, y, t) \in \mathrm{cl\, conv}(X)$, then (23) holds only when conditions (22a), (22b), (22d), (22e), and (22g) are satisfied. Conditions (22c) and (22f) do not affect the validity of (23) but if they are not satisfied then (23) is weak, i.e., a stronger inequality can be obtained from another choice of L and U. □

Remark 6 If $\bar{y}(N^+) < \bar{y}(N^-)$, then condition (22d) in Proposition 7 cannot be satisfied. However, in this case, the role of N^+ and N^- can be interchanged to satisfy (22d); interchanging N^+ and N^- is equivalent to letting $\alpha \in B^-$. □

Proof of Proposition 7 Let us define auxiliary variables $\beta, \gamma \in \mathbb{R}$ as $\beta = \max_\alpha(N^-)$ and $\gamma = \max_\alpha(S^+)$, respectively. Then, inequality (21) reduces to

$$\bar{t} \geq \max_{S^+ \subseteq N^+} \max_{\alpha, \beta, \gamma} -\frac{\gamma^2}{4} - \sum_{i \in N^+\setminus S^+} \frac{(\alpha_i^2 - \gamma^2)_+}{4} \bar{x}_i + \alpha' \bar{y} \qquad (24a)$$

$$\text{s.t. } \alpha_i \leq \gamma, \quad \forall i \in S^+ \qquad (24b)$$

$$\alpha_i \leq \beta, \quad \forall i \in N^- \qquad (24c)$$

$$\beta \leq -\alpha_i, \ \forall i \in N^+ \qquad (24d)$$

$$\alpha \in \mathbb{R}^N, \ \gamma \in \mathbb{R}_+, \beta \in \mathbb{R}_-, \qquad (24e)$$

where constraints (24b) and (24c) enforce the definitions of γ and β, and constraints (24d) and (24e) enforce that $\alpha \in B^+$.

First, observe that there exists an optimal solution of (24) with $\gamma \leq \alpha_i$ for all $i \in N^+$: if $\alpha_i < \gamma$ for some $i \in N^+$, then setting $\alpha_i = \gamma$ results in a feasible solution with improved objective value. Therefore, the value of S^+ is completely determined by γ since $S^+ = \{i \in N^+ : \alpha_i \leq \gamma\}$. Also note that $\alpha_i = \beta$ for all $i \in N^-$: if $\alpha_i < \beta$ for some $i \in N^-$, then setting $\alpha_i = \beta$ results in an improved (or identical) objective value. We now consider two cases:

Case 1 Suppose in an optimal solution of (24) we have $\gamma = -\beta$, which implies that $\alpha_i = \gamma$ for all $i \in N^+$ and $\alpha_i = -\gamma$ for all $i \in N^-$. In this case, (24) simplifies to $\bar{t} \geq \max_{\gamma \in \mathbb{R}_+} \gamma(\bar{y}(N^+) - \bar{y}(N^-)) - \frac{\gamma^2}{4}$, which, after optimizing for γ, further reduces to the original rank-one quadratic inequality $\bar{t} \geq (\bar{y}(N^+) - \bar{y}(N^-))^2$.

Case 2 Now suppose $\gamma < -\beta$ in an optimal solution. Let $L = \{i \in N^+ : \alpha_i = \gamma\}$ and $U = \{i \in N^+ : \alpha_i = -\beta\}$. Then, from the discussion above, (24) reduces to

$$t \geq \max_{\alpha, \beta, \gamma} \gamma \cdot \bar{y}(L) - \frac{\gamma^2}{4}\big(1 - \bar{x}(N^+\setminus L)\big) + \sum_{i \in N^+\setminus(L \cup U)} \left(\alpha_i \bar{y}_i - \frac{\alpha_i^2}{4} \bar{x}_i\right)$$

$$- \beta\big(\bar{y}(U) - \bar{y}(N^-)\big) - \frac{\beta^2}{4}\bar{x}(U) \qquad (25a)$$

$$\text{s.t. } \gamma < \alpha_i < -\beta, \quad \forall i \in N^+\backslash(L \cup U) \tag{25b}$$

$$\alpha \in \mathbb{R}^N, \ \gamma, \beta \in \mathbb{R}_+. \tag{25c}$$

Observe that for (L, U, γ) to correspond to an optimal solution, we must have $1 - \bar{x}(N^+\backslash L) \geq 0$ (otherwise, γ can be increased to another α_i while improving the objective value) and $\bar{y}(U) - \bar{y}(N^-) \geq 0$ (otherwise, $-\beta$ can be decreased to another α_i while improving the objective value). When both conditions are satisfied, from first-order conditions we see that $\alpha_i = 2\bar{y}_i/\bar{x}_i$ for $i \in N^+\backslash(L \cup U)$, $\gamma = 2\bar{y}(L)/\big(1 - \bar{x}(N^+\backslash L)\big)$ and $\beta = -2\big(\bar{y}(U) - \bar{y}(N^-)\big)/\bar{x}(U)$, and (25) simplifies to (23). The constraints $\gamma < \alpha_i$ are satisfied for all $i \in N^+\backslash(L \cup U)$ if and only if (22b) hold, constraints $\alpha_i \leq -\beta$ are satisfied for all $i \in N^+\backslash(L \cup U)$ if and only if (22e) hold, and constraint $\alpha < -\beta$, which may not be implied if $N^+\backslash(L \cup U) = \emptyset$, is satisfied if and only if (22g) holds.

Finally, we verify that first order conditions are satisfied for $j \in L$, this is, setting $\alpha_j > \gamma$ results in a worse solution. If condition (22c)

$$\frac{\bar{y}(L)}{1 - \bar{x}(N^+\backslash L)} \geq \frac{\bar{y}_i}{\bar{x}_i}, \quad \forall i \in L$$

does not hold for some $j \in L$, then increasing α_j from $\gamma = 2\frac{\bar{y}(L)}{1-\bar{x}(N^+\backslash L)}$ to $2\bar{y}_j/\bar{x}_j$ improves the objective value. Similarly, we verify that first order conditions for $j \in U$: if condition (22f)

$$\frac{\bar{y}(U) - \bar{y}(N^-)}{\bar{x}(U)} \leq \frac{\bar{y}_i}{\bar{x}_i}, \quad \forall i \in U$$

does not hold for some $j \in U$, then α_j can be decreased from $\beta = \frac{\bar{y}(U)-\bar{y}(N^-)}{\bar{x}(U)}$ to improve the objective value.

Note that conditions (22b) and (22c) together imply that $\bar{y}_i/\bar{x}_i < \bar{y}_j/\bar{x}_j$ whenever $i \in L$ and $j \notin L$; in other words, if $L \neq \emptyset$, then $L = \{(1), (2), \ldots, (\kappa_1)\}$ for some $1 \leq \kappa_1 \leq m$. Similarly, from conditions (22e) and (22f), we conclude that either $U = \emptyset$ or $U = \{(\kappa_2), (\kappa_2 + 1), \ldots, (m)\}$ for some $1 \leq \kappa_2 \leq m$. □

5.1.2 Lifted inequalities for X_+

We now present the inequalities for X_+, which can be interpreted as a special cases of the inequalities for X given in Sect. 5.1.1. Recall that for set X_+, the set B used in (6a) is simply $B = \mathbb{R}^N$ (we can assume $B = \mathbb{R}^N_+$ without loss of generality) and a closed form expression for (6a) requires solving the lifting problem

$$t \geq \max_{S \subseteq N} \max_{\alpha \in \mathbb{R}^N_+} -\frac{\max_\alpha(S)^2}{4} - \sum_{i \in N\backslash S} \frac{\big(\alpha_i^2 - \max_\alpha(S)^2\big)_+}{4} x_i + \alpha' y. \tag{26}$$

Note that in the proof of Proposition 7, set U corresponds to the set of variables in N^+ where constraint $\alpha_i \leq -\max_\alpha(N^-)$ is tight in an optimal solution of (24).

Intuitively, set X_+ can be interpreted as a special case of X where $N^+ = N$ and $N^- = \emptyset$, and such constraints can be dropped from the lifting problem. Therefore, we may assume $U = \emptyset$ in Proposition 7. Proposition 8 formalizes this intuition; note however that it is slightly stronger as, unlike Proposition 7, it guarantees the existence of a set satisfying the conditions of the proposition. Similar to Proposition 7, index variables in N so that $\bar{y}_{(1)}/\bar{x}_{(1)} \leq \bar{y}_{(2)}/\bar{x}_{(2)} \leq \cdots \leq \bar{y}_{(n)}/\bar{x}_{(n)}$.

Proposition 8 *Given $(\bar{x}, \bar{y}, \bar{t}) \in [0, 1]^N \times \mathbb{R}_+^N \times \mathbb{R}$, there exists an index $0 \leq \kappa \leq n$ such that the (possibly empty) set $L = \{(i) \in N : i \leq \kappa\}$ satisfies*

$$1 - \bar{x}(N \backslash L) \geq 0 \tag{27a}$$

$$\frac{\bar{y}(L)}{1 - \bar{x}(N \backslash L)} < \frac{\bar{y}_i}{\bar{x}_i}, \quad i \in N \backslash L \tag{27b}$$

$$\frac{\bar{y}(L)}{1 - \bar{x}(N \backslash L)} \geq \frac{\bar{y}_i}{\bar{x}_i}, \quad i \in L \tag{27c}$$

and inequality (26) *is satisfied if and only if*

$$t \geq \frac{\bar{y}(L)^2}{1 - \bar{x}(N \backslash L)} + \sum_{i \in N \backslash L} \frac{\bar{y}_i^2}{\bar{x}_i}. \tag{28}$$

The proof of Proposition 8 is given in "Appendix A".

Example 2 (cont) Consider X_+ with $n = 3$, and assume $x_2 = 0.6$, $x_3 = 0.3$, $y_2 = 0.5$ and $y_3 = 0.2$. Note that $y_2/x_2 \approx 0.83 > 0.67 \approx y_3/x_3$. We now compute the minimum values t such that $(x, y, t) \in \text{cl conv}(X_+)$, for different values of (x_1, y_1).

- Let $(x_1, y_1) = (0.01, 1)$ and $y_1/x_1 = 100$. Then $L = \emptyset$ satisfies all conditions (22): $x(N) = 0.91 < 1$, conditions (27b) are trivially satisfied since $y(\emptyset) = 0$, and conditions (27c) are void. In this case, we find that $(x, y, t) \in \text{cl conv}(X_+)$ iff $t \geq 1^2/0.01 + 0.5^2/0.6 + 0.2^2/0.3 \approx 100.55$. In contrast, $(x, y, t) \in \text{cl conv}(X_f)$ iff $t \geq (0.01 + 0.5 + 0.2)^2 /0.91 \approx 3.18$.
- Let $(x_1, y_1) = (0.1, 0.5)$ and $y_1/x_1 = 5$. Then $L = \{3\}$ satisfies all conditions (22): $x_1 + x_2 = 0.7 < 1$, $0.2/0.3 \approx 0.67 < y_2/x_2$ and $0.2/0.3 \approx 0.67 = y_3/x_3$. In this case, $(x, y, t) \in \text{cl conv}(X_+)$ iff $t \geq 0.2^2/0.3 + 0.5^2/0.1 + 0.5^2/0.6 \approx 3.05$. In contrast, $(x, y, t) \in \text{cl conv}(X_f)$ iff $t \geq (0.5 + 0.5 + 0.2)^2 /1 = 1.44$.
- Let $(x_1, y_1) = (0.4, 0.1)$ and $y_1/x_1 = 0.25$. Then $L = \{1, 3\}$ satisfies all conditions (22): $x_2 = 0.6 < 1$, $(0.1 + 0.2)/0.4 = 0.75 < y_2/x_2$ and $(0.1 + 0.2)/0.4 = 0.75 \geq y_3/x_3$. In this case, $(x, y, t) \in \text{cl conv}(X_+)$ iff $t \geq (0.1 + 0.2)^2/0.4 + 0.5^2/0.6 \approx 0.642$. In contrast, $(x, y, t) \in \text{cl conv}(X_f)$ iff $t \geq (0.1 + 0.5 + 0.2)^2 = 0.640$.
- Let $(x_1, y_1) = (0.5, 0.2)$ and $y_1/x_1 = 0.4$. Then $L = \{1, 2, 3\}$ satisfies all conditions (22): (27a) is trivially satisfied, (27b) is void and $(0.2 + 0.5 + 0.2)/1 = 0.9 \geq y_2/x_2$. In this case, $(x, y, t) \in \text{cl conv}(X_+)$ iff $t \geq (0.2 + 0.5 + 0.2)^2 = 0.81$, which coincides with $\text{cl conv}(X_f)$ and the natural inequality $t \geq y(N)^2$.

Figure 1 plots the minimum values of t as a function of (x_1, y_1) for $\text{cl conv}(X_f)$ and $\text{cl conv}(X_+)$. □

5.1.3 Separation

We now consider the separation problem for inequalities (21) and (26), i.e., given a point $(\bar{x}, \bar{y}, \bar{t}) \in [0, 1]^N \times \mathbb{R}_+^N \times \mathbb{R}$, finding sets $L, U \subseteq N^+$ satisfying the conditions in Proposition 7 or finding $L \subseteq N$ satisfying the conditions in Proposition 8, respectively.

Separation for (21) First, as pointed out in Remark 6, we verify whether $\bar{y}(N^+) \geq \bar{y}(N^-)$ or $\bar{y}(N^+) < \bar{y}(N^-)$; in the first case, we use directly the conditions in Proposition 7, and in the second one, we interchange the roles of N^+ and N^- so that $\bar{y}(N^+) \geq \bar{y}(N^-)$. Next, indexing the variables so that $\bar{y}_{(1)}/\bar{x}_{(1)} \leq \cdots \leq \bar{y}_{(m)}/\bar{x}_{(m)}$, where $m = |N^+|$, can be done in $O(m \log m)$ by sorting. Finally, one can simply enumerate all $m(m-1)/2$ possible values of (κ_1, κ_2) and verify whether conditions (22) are satisfied for each candidate set L and U. Hence, the separation algorithm runs in $O(n^2)$ time.

Separation for (26) First, indexing the variables so that $\bar{y}_{(1)}/\bar{x}_{(1)} \leq \cdots \leq \bar{y}_{(n)}/\bar{x}_{(n)}$ can be accomplished in $O(n \log n)$ time by sorting. Then, one can simply enumerate all n possible values of κ and verify whether conditions (27) are satisfied for each candidate set L. Since the sorting step dominates the complexity, the separation algorithm runs in $O(n \log n)$.

5.2 Conic quadratic valid inequalities in an extended formulation

Inequalities (23) and (28) given in the original space of variables are valid only over restricted parts of the domain. They are neither valid nor convex over the entire domain of the variables, e.g., (23) is not convex whenever $x(N^+\backslash L) \geq 1$. Thus, such inequalities are difficult to utilize directly by the optimization solvers. In order to address this challenge, in this section, we give valid conic quadratic reformulations in an extended space, which can be readily used by conic quadratic solvers.

For a partitioning (L, R, U) of N^+ consider the inequality

$$t \geq \min_{\lambda, \mu, \zeta} \frac{\left(y(L) - \lambda_0\right)^2}{1 - x(R) - x(U) + \mu(R) + \mu_0} + \sum_{i \in R} \frac{(y_i - \lambda_i)^2}{x_i - \mu_i} + \frac{\left(y(U) - y(N^-) + \lambda_0 + \lambda(R) + \zeta\right)^2}{x(U) - \mu_0}$$
$$\tag{29a}$$

$$\text{s.t. } 1 - x(R) - x(U) + \mu(R) + \mu_0 \geq 0 \tag{29b}$$

$$\mu_i \leq x_i, \quad i \in R \tag{29c}$$

$$\mu_0 \leq x(U) \tag{29d}$$

$$\lambda, \mu \in \mathbb{R}_+^R, \ \lambda_0, \mu_0, \zeta \in \mathbb{R}_+. \tag{29e}$$

Note that each inequality (29) requires $O(n)$ additional variables and constraints. Moreover, although not explicitly enforced, it is easy to verify that there exists an optimal solution to (29) with $\lambda_i \leq y_i$ and $\lambda_0 \leq y(L)$. Inequalities (29) are convex as they involve linear constraints and sums of ratios of convex quadratic terms and nonnegative linear terms, thus conic quadratic representable [3, 38]. We show, in Proposition 9, that inequalities (29) imply the strong formulations described in Proposition 7, and, in Proposition 10, that they are valid for X.

Proposition 9 *If conditions* (22a), (22b), (22d), (22e) *and* (22g) *are satisfied, then* $\lambda = \mu = 0$ *and* $\lambda_0 = \mu_0 = \zeta = 0$ *in an optimal solution of* (29).

Proof Observe that ζ does not appear in any constraint of (29). Thus, since $y(U) - y(N^-) \geq 0$ and $\lambda, \lambda_0 \geq 0$, it follows that $\zeta = 0$ in an optimal solution. Moreover, since (22a) is satisfied, then setting $\mu = 0$ is feasible for (29). Finally, find that KKT conditions are satisfied for $\lambda, \mu = 0$ and $\lambda_0 = \mu_0 = 0$ if

$$-\frac{y(L)}{1 - x(R) - x(U)} + \frac{y(U) - y(N^-)}{x(U)} \geq 0 \qquad (\lambda_0)$$

$$-\left(\frac{y(L)}{1 - x(R) - x(U)}\right)^2 + \left(\frac{y(U) - y(N^-)}{x(U)}\right)^2 \geq 0 \qquad (\mu_0)$$

$$-\frac{y_i}{x_i} + \frac{y(U) - y(N^-)}{x(U)} \geq 0, \quad \forall i \in R \qquad (\lambda_i)$$

$$-\left(\frac{y(L)}{1 - x(R) - x(U)}\right)^2 + \left(\frac{y_i}{x_i}\right)^2 \geq 0, \quad \forall i \in R. \qquad (\mu_i)$$

The KKT condition above for λ_0 is precisely (22g). Since $x(R) + x(U) \leq 1$ by (22a), and $y(U) - y(N^-) \geq 0$ by (22d), the KKT condition for μ_0 is equivalent to $\frac{y(L)}{1-x(R)-x(U)} + \frac{y(U)-y(N^-)}{x(U)} \geq 0$, and thus reduces to (22g). The KKT conditions for λ_i are satisfied since (22e) holds. Finally, the KKT conditions for μ_i can be equivalently stated as $\frac{y(L)}{1-x(R)-x(U)} \leq \frac{y_i}{x_i}$ (since $x(R)+x(U) \leq 1$ and $x, y \geq 0$), which are satisfied since (22b) holds. □

Note that when $\lambda = \mu = 0$ and $\lambda_0 = \mu_0 = \zeta = 0$, inequality (29) reduces to (23). Thus, if sets L, U satisfy the conditions of Proposition 7 for a given (x, y), then there exists $t \in \mathbb{R}$ such that $(x, y, t) \in \text{conv}(X)$ and (29) holds at equality. It remains to prove that inequalities (29) do not cut-off any points in X for any choice of partition (L, R, U).

Proposition 10 *For any partitioning* (L, R, U) *of* N^+, *inequalities* (29) *are valid for* X.

Proof It suffices to show that for any $(x, y) \in X$, i.e., $x_i \in \{0, 1\}$ and $x_i(1 - y_i) = 0$ for all $i \in N$, there exists $(\lambda, \mu, \lambda_0, \mu_0, \zeta)$ satisfying (29b)–(29e) such that inequality (29a) is valid. We prove the result by cases.

Case 1 $y(N^+) < y(N^-)$: In this case, we can set $\lambda_i = y_i$ and $\mu_i = x_i$ for $i \in R$, $\lambda_0 = y(L)$, $\mu_0 = x(U)$, $\zeta = y(N^-) - y(U) - y(L) - y(R)$, and inequality (29a) reduces to $t \geq 0$, which is valid.

Case 2 $y(N^+) \geq y(N^-)$, $x(R) = 0$ and $x(U) = 0$: In this case, $y_i = 0$, $i \in R \cup U$. Setting $\mu_i = \lambda_i = 0$ for $i \in R$, $\lambda_0 = y(N^-)$, $\mu_0 = 0$ and $\zeta = 0$, we find that

inequality (29a) reduces to $t \geq \left(y(L) - y(N^-)\right)^2 = \left(y(N^+) - y(N^-)\right)^2$, which is valid.

Case 3 $y(N^+) \geq y(N^-)$ and $x(U) \geq 1$: Setting $\lambda_i = y_i$ and $\mu_i = x_i$ for $i \in R$, $\lambda_0 = y(L)$, $\mu_0 = x(U) - 1$, and $\zeta = 0$, inequality (29a) reduces to $t \geq \left(y(N^+) - y(N^-)\right)^2$, which is valid.

Case 4 $y(N^+) \geq y(N^-)$, $x(U) = 0$, $x(R) \geq 1$, $y(N^-) < y_i$ for all $i \in R$ and $y(N^-) < y(L)$: In this case, $y_i = 0$, for all $i \in U$ and $x_i = 1$, for all $i \in R$, we can set $\mu_0 = 0$, and inequality (29) reduces to

$$t \geq \min_{\lambda,\mu} \frac{\left(y(L) - \lambda_0\right)^2}{1 - |R| + \mu(R)} + \sum_{i \in R} \frac{(y_i - \lambda_i)^2}{1 - \mu_i} \tag{30a}$$

$$\text{s.t. } 1 - |R| + \mu(R) \geq 0 \tag{30b}$$

$$\mu_i \leq 1 \quad \forall i \in R \tag{30c}$$

$$- y(N^-) + \lambda_0 + \lambda(R) + \zeta = 0 \tag{30d}$$

$$\lambda, \mu \in \mathbb{R}_+^R, \ \lambda_0, \zeta \in \mathbb{R}_+. \tag{30e}$$

Constraint (30d) is obtained since the denominator of the third term in (29a) is zero, thus constraining the numerator to vanish as well. Moreover, since variable $\zeta \geq 0$ only appears in (30d), after projecting ζ out we find that constraint (30d) reduces to

$$\lambda_0 + \lambda(R) \leq y(N^-). \tag{31}$$

Note that constraint (31), and assumptions $y(N^-) < y_i$ for all $i \in R$ and $y(N^-) < y(L)$, imply that $\lambda_i \leq y_i$ and $\lambda_0 \leq y(L)$. Observe that we can set

$$\mu_i = 1 - \frac{y_i - \lambda_i}{y(L) + y(R) - \lambda(R) - \lambda_0} \quad \forall i \in R.$$

Indeed, for any feasible λ, $y(L) + y(R) - \lambda(\bar{R}) - \lambda_0 \geq y(L) + y(R) - y(N^-) \geq 0$; thus $\mu_i \leq 1$. Moreover,

$$\frac{y_i - \lambda_i}{y(L) + y(R) - \lambda(R) - \lambda_0} \leq \frac{y_i - \lambda_i}{y(L) + y(R \setminus i) + y_i - \lambda_i} \leq 1$$

thus $\mu_i \geq 0$. For this choice of μ, we find that

$$1 - |R| + \mu(R) = \frac{y(L) - \lambda_0}{y(L) + y(R) - \lambda(R) - \lambda_0} \geq 0.$$

Finally, substituting $1 - |R| + \mu(R)$ and μ_i in (30a) with their respective values, (30a) reduces to

$$t \geq \min_{\lambda} \ \left(y(L) - \lambda_0\right)\left(y(L) + y(R) - \lambda(R) - \lambda_0\right)$$

$$+ \Big(y(L) + y(R) - \lambda(R) - \lambda_0 \Big) \sum_{i \in R} (y_i - \lambda_i)$$

$$\Leftrightarrow t \geq \min_{\lambda} \ \Big(y(L) + y(R) - \lambda(R) - \lambda_0 \Big)^2 = \Big(y(L) + y(R) - y(N^-) \Big)^2,$$

and since $y(L) + y(R) = y(N^+)$, this inequality is valid.

Case 5 $y(N^+) \geq y(N^-)$, $x(U) = 0$, $x(R) \geq 1$, $y(N^-) < y(L)$ but $y(N^-) \geq y_j$ for some $j \in R$: In this case, $y_i = 0$ for all $i \in U$, and we set $\mu_0 = 0$. Note that, in (29), we can set $\lambda_j = y_j$ and $\mu_j = x_j$, resulting in the inequality

$$t \geq \min_{\lambda, \mu, \zeta} \ \frac{\Big(y(L) - \lambda_0 \Big)^2}{1 - x(R \setminus j) - x(U) + \mu(R \setminus j)} + \sum_{i \in R \setminus j} \frac{(y_i - \lambda_i)^2}{x_i - \mu_i}$$

$$+ \frac{\Big(y(U) - y(N^-) + y_j + \lambda_0 + \lambda(R \setminus j) + \zeta \Big)^2}{x(U)}$$

$$\text{s.t. } 1 - x(R \setminus j) - x(U) + \mu(R \setminus j) \geq 0$$

$$\mu_i \leq x_i \quad \forall i \in R \setminus j$$

$$\lambda, \mu \in \mathbb{R}_+^{R \setminus j}, \ \lambda_0, \zeta \in \mathbb{R}_+.$$

This inequality of the same form as (29) but with $\hat{R} = R \setminus j$ and $\hat{y}(N^-) = y(N^-) - y_j$. After repeating sequentially this process so that $\lambda_i = y_i$ and $\mu_i = x_i$ for some subset $T \subseteq R$, such that $y(N^-) - y(T) \leq y_i$ for all $i \in R \setminus T$, and applying a similar strategy as in *Case 4*, we obtain an inequality of the form

$$t \geq \Big(y(L) + y(R \setminus T) - \big(y(N^-) - y(T) \big) \Big)^2 = \Big(y(N^+) - y(N^-) \Big)^2,$$

which is valid.

Case 6 $y(N^+) \geq y(N^-)$, $x(U) = 0$, $x(R) \geq 1$, and $y(N^-) \geq y(L)$: In this case, we can set $\lambda_0 = y(L)$, $\mu_0 = 0$, and (29) reduces to

$$t \geq \min_{\lambda, \mu} \ \sum_{i \in R} \frac{(y_i - \lambda_i)^2}{x_i - \mu_i}$$

$$\text{s.t. } 1 - x(R) + \mu(R) \geq 0$$

$$\mu_i \leq x_i \quad \forall i \in R$$

$$\lambda(R) \leq y(N^-) - y(L)$$

$$\lambda, \mu \in \mathbb{R}_+^R.$$

Moreover, if $y(N^-) - y(L) \geq y_j$ for some $j \in R$, then we can set $\lambda_j = y_j$, $\mu_j = y_j$ as done in *Case 5*. After repeating this process, we obtain an inequality of the form

$$t \geq \min_{\lambda,\mu} \sum_{i \in R\setminus T} \frac{(y_i - \lambda_i)^2}{x_i - \mu_i} \tag{32a}$$

$$\text{s.t. } 1 - x(R\setminus T) + \mu(R\setminus T) \geq 0 \tag{32b}$$

$$\mu_i \leq x_i \quad \forall i \in R\setminus T \tag{32c}$$

$$\lambda(R\setminus T) \leq y(N^-) - y(L) - y(T) \tag{32d}$$

$$\lambda, \mu \in \mathbb{R}_+^{R\setminus T}, \tag{32e}$$

where $y(N^-) - y(L) - y(T) < y_i$ for all $i \in R\setminus T$, and therefore $x_i = 1$ for all $i \in R\setminus T$.

Note that constraint (32d) and $y(N^-) - y(L) - y(T) < y_i$ imply that $\lambda_i < y_i$ in any feasible solution. Then, for all $i \in R\setminus T$, we can set

$$\mu_i = x_i - \frac{y_i - \lambda_i}{y(R\setminus T) - \lambda(R\setminus T)}.$$

Clearly, $\mu_i \leq x_i$. Moreover, for all $i \in R\setminus T$,

$$\frac{y_i - \lambda_i}{y(R\setminus T) - \lambda(R\setminus T)} \leq \frac{y_i - \lambda_i}{y(R\setminus(T \cup i)) + y_i - \lambda_i} \leq 1 = x_i,$$

thus $\mu_i \geq 0$. Finally,

$$1 - x(R\setminus\Lambda) + \mu(R\setminus T) = 1 - \frac{y(R\setminus T) - \lambda(R\setminus\lambda)}{y(R\setminus T) - \lambda(R\setminus T)} = 0,$$

and constraint (32b) is satisfied. Substituting $x_i - \lambda_i$, $i \in R\setminus T$, with their explicit form in (32a), we find the equivalent form

$$t \geq \min_{\lambda} \left(y(R\setminus T) - \lambda(R\setminus T)\right) \sum_{i \in R\setminus T} (y_i - \lambda_i) = \min_{\lambda} \left(y(R\setminus T) - \lambda(R\setminus T)\right)^2$$

$$= \left(y(N^+) - y(N^-)\right)^2,$$

which is valid. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

To derive the corresponding lifted inequalities for B^-, it suffices to interchange N^+ and N^-. Therefore, for a partitioning (L, R, U) of N^-, we find the conic quadratic inequalities:

$$t \geq \min_{\lambda,\mu,\zeta} \frac{(y(L) - \lambda_0)^2}{1 - x(R) + x(U) + \mu(R) + \mu_0} + \sum_{i \in R} \frac{(y_i - \lambda_i)^2}{x_i - \mu_i} + \frac{\left(y(U) - y(N^+) + \lambda_0 + \lambda(R) + \zeta\right)^2}{x(U) - \mu_0} \tag{33a}$$

$$\text{s.t. } 1 - x(R) - x(U) + \mu(R) + \mu_0 \geq 0 \tag{33b}$$

$$\mu_i \leq x_i, \quad i \in R \tag{33c}$$

$$\mu_0 \leq x(U) \tag{33d}$$

$$\lambda, \mu \in \mathbb{R}_+^R, \ \lambda_0, \mu_0, \zeta \in \mathbb{R}_+. \tag{33e}$$

One of the main results of the paper, that is, an explicit description of cl conv(X) via a finite number of conic quadratic inequalities, is stated below.

Theorem 3 *cl conv(X) is given by bound constraints $0 \leq x \leq 1$, $y \geq 0$, and inequalities* (29) *and* (33).

For the positive case of X_+ with $N^- = \emptyset$, for a partitioning (L, R) of N, inequalities (29) reduce to

$$t \geq \min_{\mu} \frac{y(L)^2}{1 - x(R) + \mu(R)} + \sum_{i \in R} \frac{y_i^2}{x_i - \mu_i} \tag{34a}$$

$$\text{s.t. } 1 - x(R) + \mu(R) \geq 0 \tag{34b}$$

$$\mu_i \leq x_i, \qquad i \in R \tag{34c}$$

$$\mu \in \mathbb{R}_+^R. \tag{34d}$$

Note that each inequality (34) also requires $O(n)$ additional variables and constraints but is significantly simpler compared to (29).

Theorem 4 *cl conv(X_+) is given by bound constraints $0 \leq x \leq 1$, $y \geq 0$, and inequalities* (34).

6 Computational experiments

In this section, we test the computational effectiveness of the conic quadratic inequalities given in Sect. 5.2 in solving convex quadratic minimization problems with indicators. In particular, we solve portfolio optimization problems with fixed-charges. All experiments are run with CPLEX 12.8 solver on a laptop with a 1.80GHz Intel®Core™ i7 CPU and 16 GB main memory on a single thread. We use CPLEX default settings but turn on the numerical emphasis parameter, unless stated otherwise. The data for the instances and problem formulations in `.lp` format can be found online at https://sites.google.com/usc.edu/gomez/data.

6.1 Instances

We consider optimization problems of the form

$$\min_{y,x} y'(FF')y + \sum_{i=1}^{n} (d_i y_i)^2 \tag{35a}$$

$$\text{s.t. } e'y = 1 \tag{35b}$$

$$b'y - a'x \geq \beta \tag{35c}$$

$$y_i \leq x_i, \quad i \in N \tag{35d}$$

$$x \in \{0, 1\}^N, \, y \in \mathbb{R}_+^N \tag{35e}$$

where $F \in \mathbb{R}_+^{n \times r}$ with $r < n$, $a, b, d \in \mathbb{R}_+^N$. We test two classes of instances, general and positive, where either F has both positive and negative entries, or F has only non-negative entries, respectively. Note that constraints (35d) are in fact a big-M reformulation of complementarity constraint $y_i(1 - x_i) = 0$: indeed, constraint (35b) and $y \geq 0$ imply the upper bound $y \leq 1$. The parameters are generated as follows—we use the notation $Y \sim U[\ell, u]$ as "Y is generated from a continuous uniform distribution between ℓ and u":

F Let ρ be a positive weight parameter. Matrix $F = EG$ where $E \in \mathbb{R}_+^{n \times r}$ is an exposure matrix such that $E_{ij} = 0$ with probability 0.8 and $E_{ij} \sim U[0, 1]$ otherwise, and $G \in \mathbb{R}_+^{r \times r}$ such that: $G_{ij} \sim U[\rho, 1]$. If $\rho \geq 0$, then matrix F is guaranteed to be positive, and we refer to such instances as positive. Otherwise, for $\rho < 0$, we refer to the instances as general.

d Let δ be a diagonal dominance parameter. Define $v = (1/n) \sum_{i=1}^n (FF')_{ii}$ to be the average diagonal element of FF'; then $d_i^2 \sim U[0, v\delta]$.

b We generate entries $b_i \sim U[0.25, 0.75] \times \sqrt{(FF')_{ii} + d_i^2}$. Note that if the terms b_i and $((FF')_{ii} + d_i^2)$ are interpreted as the expectation and variance of a random variable, then expectations are approximately proportional to the standard deviations. This relation aims to avoid trivial instances, where one term dominates the other.

a Let ω be a fixed cost parameter and $a_i = \omega(e'b)/n$, $i \in N$, where e is an n-dimensional vector of ones.

It is well-documented in the literature that for matrices with large diagonal dominance the perspective reformulation achieves close to 100% gap improvement. Therefore, we choose a low diagonal dominance $\delta = 0.01$ to generate instances hard for the perspective reformulation. In our computations, unless stated otherwise, we use $n = 200$ and $\beta = (e'b)/n$.

6.2 Methods

We test the following methods:

- Basic : Problem (35) formulated as

$$\min \|q\|_2^2 + \sum_{i=1}^n (d_i y_i)^2 \tag{36a}$$

$$\text{s.t. } q = F'y \tag{36b}$$

$$(35b) - (35d) \tag{36c}$$

$$x \in \{0, 1\}^n, \, y \in \mathbb{R}_+^n, \, q \in \mathbb{R}^r. \tag{36d}$$

- Perspective : Problem (35) formulated as

$$\min \|q\|_2^2 + \sum_{i=1}^n d_i^2 p_i \tag{37a}$$

$$\text{s.t. } q = F'y \tag{37b}$$

$$y_i^2 \le p_i x_i, \qquad\qquad i = 1, \dots, n \tag{37c}$$

$$(35b) - (35d) \tag{37d}$$

$$x \in \{0, 1\}^n, \ y \in \mathbb{R}_+^n, \qquad\qquad p \in \mathbb{R}_+^n, \ q \in \mathbb{R}^r. \tag{37e}$$

- Supermodular : Problem (35) formulated as

$$\min \sum_{j=1}^r t_j + \sum_{i=1}^n d_i^2 p_i \tag{38a}$$

$$\text{s.t. } \left(F_j'y\right)^2 \le t_j, \qquad\qquad j = 1, \dots, r \tag{38b}$$

$$y_i^2 \le p_i x_i, \qquad\qquad i = 1, \dots, n \tag{38c}$$

$$(35b) - (35d) \tag{38d}$$

$$x \in \{0, 1\}^n, \ y \in \mathbb{R}_+^n, \ p \in \mathbb{R}_+^n, \qquad t \in \mathbb{R}_+^r, \tag{38e}$$

where F_j denotes the j-th column of F. Additionally, lifted supermodular inequalities (29) are added to strengthen the relaxations. Note that the convex relaxation of (38) without any additional inequalities is equivalent to the convex relaxation of (37).

Cuts (29) (for general instances) or (34) (for positive instances) for method Supermodular are added as follows:

(1) We solve the convex relaxation of (38) to obtain a solution $(\bar{x}, \bar{y}, \bar{t})$. By default, the convex relaxation is solved with an interior point method.
(2) We find a most violated inequality (29) or (34) for each constraint (38b) using the separation algorithm given in Sect. 5.1.3. Denote by $\bar{v}_j$ the rhs value of (23) or (28) if sets L and U satisfying (22) exist; otherwise, let $\bar{v} = -\infty$.
(3) Let $\epsilon = 10^{-3}$ be a precision parameter. Inequalities found in step (2) are added if either $\bar{t}_j < \epsilon$ and $(\bar{v}_j - \bar{t}_j) > \epsilon$; or $\bar{t}_j \ge \epsilon$ and $(\bar{v}_j - \bar{t}_j)/\bar{t}_j > \epsilon$. At most r inequalities are added per iteration, one for each constraint (38b).
(4) This process is repeated until either no inequality is added in step (3) or max number of cuts $(3r)$ is reached.

We point out that convexification based on X_f [7], described in Sect. 3.2, is not effective with formulation (38) since $t_j \ge (F_j'y)^2 / \min\{1, e'x\}$ reduces to $t_j \ge (F_j'y)^2$ due to (35b) and (35d).

6.3 Results

Tables 1, 2, 3 and 4 present the results for $\rho = \{-1, -0.5, -0.2, 0\}$. They show, for different ranks r and values of the fixed cost parameter ω, the optimal objective value (opt) and, for each method, the optimal objective value for the convex relaxation (val), the integrality gap (gap) computed as $\text{gap} = \frac{\text{opt}-\text{val}}{\text{opt}} \times 100$, the improvement (imp) of Supermodular over Perspective computed as

$$\text{imp} = \frac{\text{gap}_{\text{Persp.}} - \text{gap}_{\text{Supermod.}}}{\text{gap}_{\text{Persp.}}},$$

the time required to solve the relaxation in seconds (time) and the number of cuts added (cuts). The optimal solutions are computed using CPLEX branch-and-bound method using the Perspective formulation. The values opt and val are scaled so that, in a given instance, opt = 100. Each row corresponds to the average of five instances generated with the same parameters.

First note that Perspective achieves only a very modest improvement over Basic due to the low diagonal dominance parameter $\delta = 0.01$. We also point out that instances with smaller positive weight ρ have weaker natural convex relaxations, i.e., Basic has larger gaps—a similar phenomenon was observed in [26].

The relative performance of all methods in rank-one instances, $r = 1$, is virtually identical regardless of the value of the positive weight parameter ρ. In particular Supermodular substantially improves upon Basic and Perspective : it achieves 0% gaps in instances with $\omega \leq 10$, and reduces to gap from 35 to 6% in instances with $\omega = 50$.

In instances with $r \geq 5$, the relative performance of Supermodular depends on the positive weight parameter ρ: for larger values of ρ, more cuts are added and Supermodular results in higher quality formulations. For example, in instances with $r = 5$, $\omega = 50$, the improvements achieved by Supermodular are 40.3% ($\rho = -1$), 53.2% ($\rho = -0.5$), 62.0% ($\rho = -0.2$) and 72.7% ($\rho = 0$). Similar behavior can be observed for other combinations of parameters with $r \geq 5$.

Our interpretation of the dependence of ρ in the strength of the formulation is as follows. For instances with small values of ρ, it is possible to reduce the systematic risk of the portfolio $y'(FF')y$ close to zero due to negative correlations, i.e., achieve "perfect hedge" although it may be unrealistic in practice. In such instances, the idiosynctratic risk $\sum_{i=1}^{n}(d_i y_i)^2$ and constraints (35b)–(35d), which limit diversification, are the most important components behind the portfolio variance. In contrast, as ρ increases, it is increasingly difficult to reduce the systematic risk (and altogether impossible for $\rho \geq 0$). Thus, in such instances, the systematic risk $y'(FF')y$ accounts for the majority of the variance of the portfolio. Thus, the lifted supermodular inequalities, which exploit the structure induced by the systematic risk, are particularly effective in the later class of instances.

Table 1 Computational results for `general` instances, $\rho = -1$

r	ω	Opt	Method	Strength			Performance	
				Val	Gap (%)	Imp (%)	Time (s)	Cuts
1	2	100.0	Basic	92.5	7.5		<0.1	–
			Perspective	98.4	1.6		<0.1	–
			Supermodular	100.0	0.0	100.0	0.1	1
	10	100.0	Basic	82.9	17.1		<0.1	–
			Perspective	90.9	9.1		<0.1	–
			Supermodular	100.0	0.0	100.0	0.1	1
	50	100.0	Basic	61.7	38.3		<0.1	–
			Perspective	65.4	34.6		<0.1	–
			Supermodular	94.3	5.7	83.5	0.1	1
5	2	100.0	Basic	88.3	11.7		<0.1	–
			Perspective	96.5	3.5		<0.1	–
			Supermodular	97.7	2.3	34.3	0.1	3
	10	100.0	Basic	69.7	30.3		<0.1	–
			Perspective	80.5	19.5		<0.1	–
			Supermodular	88.5	11.5	41.0	0.3	4
	50	100.0	Basic	41.8	58.2		<0.1	–
			Perspective	46.6	53.4		<0.1	–
			Supermodular	68.1	31.9	40.3	0.6	5
10	2	100.0	Basic	87.1	12.9		<0.1	–
			Perspective	95.6	4.4		<0.1	–
			Supermodular	95.8	4.2	4.5	0.1	2
	10	100.0	Basic	62.0	38.0		<0.1	–
			Perspective	72.9	27.1		<0.1	–
			Supermodular	76.1	23.9	11.8	0.7	7
	50	100.0	Basic	27.4	72.6		<0.1	–
			Perspective	30.9	69.1		<0.1	–
			Supermodular	40.4	59.6	13.7	1.0	12

Figure 4 depicts the integrality gap of different formulations as a function of rank for instances with $\rho = 0$. We see that `Supermodular` achieves large ($> 70\%$) improvement over `Perspective` especially in the challenging low-rank settings. The improvement is significant (44%) also for high-rank settings with $r = 35$.

Finally, to evaluate the computational burden associated with the formulations, we plot in Fig. 5 the time in seconds (in a logarithmic scale) require the solve the convex relaxations of each method for different dimensions n. Each point in Fig. 5 corresponds to an average of 15 portfolio optimization instances generated with parameters $r = 10$, $\delta = 0.01$ and $\omega \in \{2, 10, 50\}$ (5 instances for each value of ω). The time for `Supermodular` includes the total time used to generate cuts and solving the convex relaxations many times.

Table 2 Computational results for `general` instances, $\rho = -0.5$

r	ω	Opt	Method	Strength			Performance	
				Val	Gap (%)	Imp (%)	Time (s)	Cuts
1	2	100.0	Basic	92.5	7.5		<0.1	–
			Perspective	98.3	1.7		<0.1	–
			Supermodular	100.0	0.0	100.0	0.1	1
	10	100.0	Basic	82.9	17.1		<0.1	–
			Perspective	91.0	9.1		<0.1	–
			Supermodular	99.9	0.1	98.9	0.1	1
	50	100.0	Basic	61.7	38.3		<0.1	–
			Perspective	65.3	34.7		<0.1	–
			Supermodular	94.2	5.8	83.3	0.1	1
5	2	100.0	Basic	91.5	8.5		<0.1	–
			Perspective	96.5	3.5		<0.1	–
			Supermodular	98.1	1.9	45.7	0.3	4
	10	100.0	Basic	76.4	23.6		<0.1	–
			Perspective	83.1	16.9		0.1	–
			Supermodular	92.7	7.3	56.8	0.4	4
	50	100.0	Basic	52.1	47.9		<0.1	–
			Perspective	55.1	44.9		<0.1	–
			Supermodular	79.0	21.0	53.2	0.4	4
10	2	100.0	Basic	89.7	10.3		<0.1	–
			Perspective	93.3	6.7		0.1	–
			Supermodular	94.7	5.3	20.9	1.6	10
	10	100.0	Basic	69.5	30.5		<0.1	–
			Perspective	73.2	26.8		<0.1	–
			Supermodular	81.4	18.6	30.6	2.4	11
	50	100.0	Basic	38.3	61.7		<0.1	–
			Perspective	39.6	60.4		<0.1	–
			Supermodular	54.5	45.5	24.7	2.2	16

We see that, in general, formulation `Basic` is an order-of-magnitude faster than `Perspective`, which in turn is an order-of-magnitude faster than `Supermodular`. Nonetheless, the computation times for `Supermodular` are adequate for many applications, solving instances with $n = 1000$ on average under four seconds.

Contrary to expectations, `Supermodular` is faster for `general` instances than for `positive` instances, despite the larger and more complex inequalities (29) used for the `general` case; for $n = 1000$, `Supermodular` runs in 1.9 s in `general` instances versus 3.8 s in `positive` instances. This counter-intuitive behavior is explained by the number of cuts added, as several more violated cuts are found in instances with large values of ρ, leading to larger convex formulations and the need

Table 3 Computational results for `general` instances, $\rho = -0.2$

r	ω	Opt	Method	Strength			Performance	
				Val	Gap (%)	Imp (%)	Time (s)	Cuts
1	2	100.0	Basic	92.5	7.5		<0.1	–
			Perspective	98.3	1.7		<0.1	–
			Supermodular	100.0	0.0	100.0	0.1	1
	10	100.0	Basic	82.9	17.1		<0.1	–
			Perspective	90.9	9.1		<0.1	–
			Supermodular	99.9	0.1	98.9	0.1	1
	50	100.0	Basic	61.7	38.3		<0.1	–
			Perspective	65.4	34.6		<0.1	–
			Supermodular	94.2	5.8	83.2	0.1	1
5	2	100.0	Basic	93.8	6.2		<0.1	–
			Perspective	96.3	3.7		<0.1	–
			Supermodular	98.9	1.1	70.3	0.6	5
	10	100.0	Basic	79.6	20.4		<0.1	–
			Perspective	82.1	17.9		0.1	–
			Supermodular	93.8	6.2	65.4	0.8	6
	50	100.0	Basic	57.6	42.4		<0.1	–
			Perspective	58.7	41.3		<0.1	–
			Supermodular	84.3	15.7	62.0	0.6	5
10	2	100.0	Basic	93.1	6.9		<0.1	–
			Perspective	94.9	5.1		<0.1	–
			Supermodular	97.6	2.4	52.9	6.6	14
	10	100.0	Basic	77.8	22.2		<0.1	–
			Perspective	79.3	20.7		<0.1	–
			Supermodular	89.7	10.3	50.2	3.0	12
	50	100.0	Basic	56.9	43.1		<0.1	–
			Perspective	57.5	42.5		<0.1	–
			Supermodular	76.9	23.1	45.6	10.3	20

to resolve them more times; for $n = 1000$, 20 cuts are added in each instance with $\rho = 0$, whereas on average only 3.7 cuts are added in instances with $\rho = -1$.

The computation times are especially promising for tackling large-scale quadratic optimization problems with indicators, where alternatives to constructing strong convex relaxations (often based on decomposition of matrix $FF' + D$ into lower-dimensional terms) may not scale. For example, Frangioni et al. [26] solve convex relaxations of instances up to $n = 50$, Han et al. [33] solve relaxations for instances up to $n = 150$, and Atamtürk and Gómez [6] report that solving the convex relaxation of quadratic instances with $n = 200$ requires up to 1000 s. All of these methods require adding $O(n^2)$ variables and constraints to the formulations to achieve strengthening. In contrast, the supermodular inequalities (29) and (34) yield formulations with $O(nr)$

Table 4 Computational results for `positive` instances, $\rho = 0$

r	ω	Opt	Method	Strength			Performance	
				Val	Gap (%)	Imp (%)	Time (s)	Cuts
1	2	100.0	Basic	92.5	7.5		<0.1	–
			Perspective	98.3	1.7		<0.1	–
			Supermodular	100.0	0.0	100.0	0.1	2
	10	100.0	Basic	82.9	17.1		<0.1	–
			Perspective	91.0	9.0		<0.1	–
			Supermodular	99.9	0.1	98.9	0.1	2
	50	100.0	Basic	61.7	38.3		<0.1	–
			Perspective	65.3	34.7		<0.1	–
			Supermodular	94.2	5.8	83.3	0.1	2
5	2	100.0	Basic	94.1	5.9		<0.1	–
			Perspective	96.2	3.8		<0.1	–
			Supermodular	98.7	1.3	65.8	0.2	10
	10	100.0	Basic	80.4	19.6		<0.1	–
			Perspective	82.4	17.6		<0.1	–
			Supermodular	93.4	6.6	65.2	0.2	10
	50	100.0	Basic	65.6	34.4		<0.1	–
			Perspective	66.7	33.3		<0.1	–
			Supermodular	90.9	9.1	72.7	0.2	10
10	2	100.0	Basic	94.0	6.0		<0.1	–
			Perspective	95.5	4.5		<0.1	-
			Supermodular	97.6	2.4	51.1	0.6	20
	10	100.0	Basic	83.1	16.9		<0.1	–
			Perspective	84.4	15.6		<0.1	–
			Supermodular	93.2	6.8	56.4	0.6	20
	50	100.0	Basic	66.0	34.0		<0.1	–
			Perspective	66.7	33.3		<0.1	–
			Supermodular	82.6	17.4	47.7	0.6	20

additional variables and constraints, which can be solved efficiently even if n is large provided that the rank r is sufficiently small: in our computations, instances with $r = 10$ and $n = 1000$ are solved in under 4 s. Nonetheless, as discussed in the next section, even if the convex relaxations can be solved easily, incorporating the proposed convexification in branch-and-bound methods may required tailored implementations, not supported by current off-the-shelf branch-and-bound solvers.

6.4 On the performance with off-the-shelf branch-and-bound solvers

We also experimented with solving the formulations `Supermodular` obtained after adding cuts with CPLEX branch-and-bound algorithm. However, note that inequalities
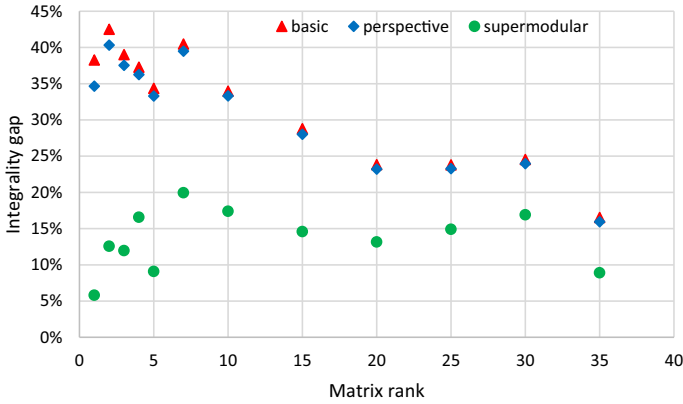
Fig. 4 Integrality gap vs matrix rank ($\omega = 50$, $\delta = 0.01$, $\rho = 0$)

(A) General instances with $\rho = -1$.

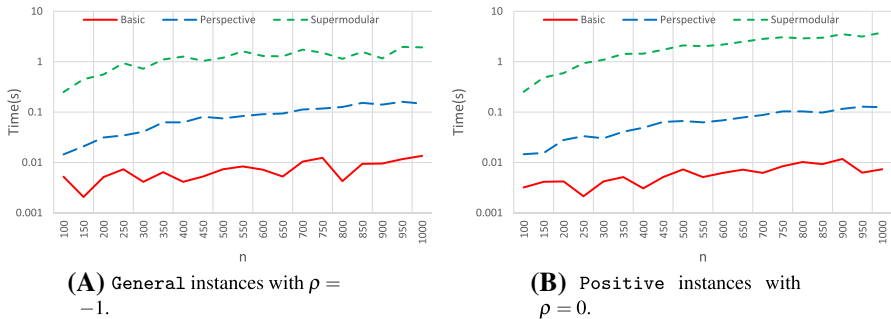**(B)** Positive instances with $\rho = 0$.

Fig. 5 Solution time vs problem dimension ($r = 10$, $\delta = 0.01$)

(29) and, to a lesser degree, inequalities (34), involve several ratios that can result in division by 0—from the proof of Proposition 10, we see that this in fact the case in many scenarios. Therefore, while we did not observe any particular numerical difficulties when solving the convex relaxations (via interior point methods), in a small subset of the instances we observed that the branch-and-bound method (based on linear outer approximations) resulted in numerical issues leading to incorrect solutions.

Table 5 reports the results on the two instances that exhibit such pathological behavior. It shows, for each instance and method and different CPLEX settings, the bounds on the optimal solution obtained reported by CPLEX when solving the convex relaxation via interior point methods (barrier, corresponding to a lower bound), and lower and upper bounds reported by running the branch-and-bound algorithm for one hour. We do not scale the solutions obtained in Table 5. The tested settings are default CPLEX (def), default CPLEX with numerical emphasis enabled (+num), and CPLEX with numerical emphasis enabled and presolve and CPLEX cuts disabled (+num-pc).

In the first instance shown in Table 5, when using Supermodular with the default CPLEX settings, the solution reported is worse than the optimal solution by 30%. By enabling the numerical emphasis option, the solution improves but is still 10% worse than the solution reported by Perspective. Nonetheless, if presolve

Table 5 Examples of pathological behavior in branch-and-bound

Instance	Method	Setting	Bounds		
			Barrier	lb_bb	ub_bb
	Perspective	def	0.0202	0.0942	0.0942
200-10-1.0-0.01	Supermodular	def	0.0243	0.1249	0.1249
−50.0-1-1-103[†]		+num	0.0243	0.1078	0.1078
		+num-pc	0.0243	0.0942	0.0942
	Perspective	def	0.1950	0.4849	0.4849
200-10-1.0-0	Supermodular	def	0.2471	0.4849	0.4849
−50.0-1-1-104[††]		+num	0.2471	0.4849	0.4849
		+num-pc	0.2471	0.5209	0.6629

† General portfolio instance with $\rho = -1$, $n = 200$, $r = 10$, $\delta = 0.01$, $\omega = 50$
†† General portfolio instance with $\rho = -1$, $n = 200$, $r = 10$, $\delta = 0$, $\omega = 50$

and CPLEX cuts are disabled, then both solutions coincide. The second instance shown in Table 5 exhibits the opposite behavior: when used with the default settings, independently of the numerical emphasis, the solutions obtained by Perspective and Supermodular coincide; however, if presolve and CPLEX cuts are disabled, then the lower bound obtained after one hour of branch-and-bound with the Supermodular method already precludes finding the correct solution. We point out that pathological behavior of conic quadratic branch-and-bound solvers have been observed in the past for other nonlinear mixed-integer problems with a large number of variables, see for example [6, 13, 26, 29].

7 Conclusions

In this paper we describe the convex hull of the epigraph of a rank-one quadratic functions with indicator variables. In order to do so, we first describe the convex hull of a underlying supermodular set function in a lower-dimensional space, and then maximally lift the resulting facets into nonlinear inequalities in the original space of variables. The approach is broadly applicable, as most of the existing results concerning convexifications of convex quadratic functions with indicator variables can be obtained in this way, as well as several well-known classes of facet-defining inequalities for mixed-integer linear problems.

Appendix A

Proof of Proposition 8 In order to solve problem (26) we introduce an auxiliary variable $\gamma \in \mathbb{R}_+$ such that $\gamma = \max_\alpha(S^+)$. Then, inequality (26) reduces to

$$\bar{t} \geq \max_{S \subseteq N} \max_{\alpha, \gamma} -\frac{\gamma^2}{4} - \sum_{i \in N \setminus S} \frac{(\alpha_i^2 - \gamma^2)_+}{4} \bar{x}_i + \alpha' \bar{y} \tag{39a}$$

$$\text{s.t. } \alpha_i \leq \gamma, \quad i \in S \tag{39b}$$

$$\alpha \in \mathbb{R}_+^N, \; \gamma \in \mathbb{R}_+, \tag{39c}$$

where constraint (39b) enforces the definition of γ.

Note that there exists an optimal solution for (39) were $\gamma \leq \alpha_i$ for all $i \in N$: if $\alpha_i < \gamma$ for some $i \in N$, then setting $\alpha_i = \gamma$ yields a feasible solution with improved objective value. Therefore, S is completely determined by γ since $S = \{i \in N : \alpha_i \leq \gamma\}$.

Now, let $L = \{i \in N : \alpha_i = \gamma\}$ in a solution of (39). From the discussion above, we find that (39) reduces to

$$t \geq \max_{\alpha, \gamma} \gamma \cdot \bar{y}(L) - \frac{\gamma^2}{4}\big(1 - \bar{x}(N \setminus L)\big) + \sum_{i \in N \setminus L} \left(\alpha_i \bar{y}_i - \frac{\alpha_i^2}{4} \bar{x}_i\right) \tag{40a}$$

$$\text{s.t. } \gamma < \alpha_i \quad \forall i \in N \setminus L \tag{40b}$$

$$\alpha \in \mathbb{R}^N, \; \gamma \in \mathbb{R}_+. \tag{40c}$$

Observe that for (L, γ) to correspond to an optimal solution, we require that $1 - \bar{x}(N \setminus L) \geq 0$ (otherwise γ can be increased and set to an upper bound while improving the objective value). When this condition is satisfied, we find by taking derivatives of the objective and setting to 0, that $\alpha_i = 2\bar{y}_i / \bar{x}_i$ for $i \in N \setminus L$ and $\gamma = 2\bar{y}(L) / \big(1 - \bar{x}(N \setminus L)\big)$, and (40) simplifies to (28). Note however that, in general, (α, γ) may not satisfy constraints (40b) for any choice of sets $L \subseteq N$. The constraints are satisfied if and only if $\gamma < \alpha_i$ for all $i \in N \setminus L$, i.e., if and only if conditions (27b) are satisfied.

In order for L to be optimal we require condition (27c), i.e.,

$$\frac{y(L)}{1 - \bar{x}(N \setminus L)} \geq \frac{\bar{y}_i}{\bar{x}_i} \quad \forall i \in L.$$

Indeed, if this condition is not satisfied for some $j \in L$, then increasing α_j from $\gamma = 2\frac{\bar{y}(L)}{1-\bar{x}(N \setminus L)} \geq 2\frac{\bar{y}_j}{\bar{x}_j}$ to $2\bar{y}_j/\bar{x}_j$ (or setting it to β if $\beta < 2\bar{y}_j/\bar{x}_j$) results in a better objective value.

Finally, note that conditions (27b) and (27c) together imply that $\bar{y}_i/\bar{x}_i < \bar{y}_j/\bar{x}_j$ whenever $i \in L$ and $j \notin L$; in other words, if $L \neq \emptyset$, then $L = \{(1), (2), \ldots, (\kappa)\}$ for some $1 \leq \kappa \leq n$. □

References

1. Ahmed, S., Atamtürk, A.: Maximizing a class of submodular utility functions. Math. Program. **128**(1–2), 149–169 (2011)
2. Aktürk, M.S., Atamtürk, A., Gürel, S.: A strong conic quadratic reformulation for machine-job assignment with controllable processing times. Oper. Res. Lett. **37**, 187–191 (2009)
3. Alizadeh, F., Goldfarb, D.: Second-order cone programming. Math. Program. **95**, 3–51 (2003)
4. Atamtürk, A.: Flow pack facets of the single node fixed-charge flow polytope. Oper. Res. Lett. **29**, 107–114 (2001)
5. Atamtürk, A., Bhardwaj, A.: Supermodular covering knapsack polytope. Discret. Optim. **18**, 74–86 (2015)
6. Atamtürk, A., Gómez, A.: Strong formulations for quadratic optimization with M-matrices and indicator variables. Math. Program. **170**, 141–176 (2018)
7. Atamtürk, A., Gómez, A.: Rank-one convexification for sparse regression (2019). arXiv:1901.10334
8. Atamtürk, A., Gómez, A.: Submodularity in conic quadratic mixed 0–1 optimization. Oper. Res. **68**(2), 609–630 (2020)
9. Atamtürk, A., Muñoz, J.C.: A study of the lot-sizing polytope. Math. Program. **99**, 443–465 (2004)
10. Atamtürk, A., Narayanan, V.: Submodular function minimization and polarity. Math. Program. **196**, 57–67 (2022)
11. Atamtürk, A., Nemhauser, G.L., Savelsbergh, M.W.P.: Valid inequalities for problems with additive variable upper bounds. Math. Program. **91**, 145–162 (2001)
12. Atamtürk, A., Küçükyavuz, S., Tezel, B.: Path cover and path pack inequalities for the capacitated fixed-charge network flow problem. SIAM J. Optim. **27**(3), 1943–1976 (2017)
13. Atamtürk, A., Gómez, A., Han, S.: Sparse and smooth signal estimation: convexification of ℓ_0-formulations. J. Mach. Learn. Res. **22**, 1–4 (2021)
14. Bach, F.: Submodular functions: from discrete to continuous domains. Math. Program. **175**, 419–459 (2019)
15. Bertsimas, D., King, A.: Or forum—an algorithmic approach to linear regression. Oper. Res. **64**, 2–16 (2015)
16. Bienstock, D.: Computational study of a family of mixed-integer quadratic programming problems. Math. Program. **74**(2), 121–140 (1996)
17. Bienstock, D., Michalka, A.: Cutting-planes for optimization of convex functions over nonconvex sets. SIAM J. Optim. **24**, 643–677 (2014)
18. Bonami, P., Lodi, A., Tramontani, A., Wiese, S.: On mathematical programming with indicator constraints. Math. Program. **151**, 191–223 (2015)
19. Ceria, S., Soares, J.: Convex programming for disjunctive convex optimization. Math. Program. **86**, 595–614 (1999)
20. Cozad, A., Sahinidis, N.V., Miller, D.C.: A combined first-principles and data-driven approach to model building. Comput. Chem. Eng. **73**, 116–127 (2015)
21. Dong, H., Linderoth, J.: On valid inequalities for quadratic programming with continuous variables and binary indicators. In: Goemans, M., Correa, J. (eds.) Proceedings of IPCO 2013, pp. 169–180. Springer, Berlin (2013)
22. Dong, H., Chen, K., Linderoth, J.: Regularization vs. relaxation: a conic optimization perspective of statistical variable selection (2015). arXiv:1510.06083
23. Frangioni, A., Gentile, C.: Perspective cuts for a class of convex 0–1 mixed integer programs. Math. Program. **106**, 225–236 (2006)

24. Frangioni, A., Gentile, C.: SDP diagonalizations and perspective cuts for a class of nonseparable MIQP. Oper. Res. Lett. **35**, 181–185 (2007)
25. Frangioni, A., Gentile, C., Lacalandra, F.: Tighter approximated MILP formulations for unit commitment problems. IEEE Trans. Power Syst. **24**(1), 105–113 (2009)
26. Frangioni, A., Gentile, C., Hungerford, J.: Decompositions of semidefinite matrices and the perspective reformulation of nonseparable quadratic programs. Math. Oper. Res. **45**(1), 15–33 (2020)
27. Fujishige, S.: Submodular Functions and Optimization, vol. 58. Elsevier, Amsterdam (2005)
28. Gómez, A.: Submodularity and valid inequalities in nonlinear optimization with indicator variables (2018)
29. Gómez, A.: Strong formulations for conic quadratic optimization with indicator variables. Math. Program. **188**, 193–226 (2020)
30. Gómez, A.: Outlier detection in time series via mixed-integer conic quadratic optimization. SIAM J. Optim. **31**, 1897–1925 (2021)
31. Grötschel, M., Lovász, L., Schrijver, A.: The ellipsoid method and its consequences in combinatorial optimization. Combinatorica **1**, 169–197 (1981)
32. Günlük, O., Linderoth, J.: Perspective reformulations of mixed integer nonlinear programs with indicator variables. Math. Program. **124**, 183–205 (2010)
33. Han, S., Gómez, A., Atamtürk, A.: 2x2 convexifications for convex quadratic optimization with indicator variables (2020). arXiv:2004.07448
34. Hazimeh, H., Mazumder, R., Saab, A.: Sparse regression at scale: branch-and-bound rooted in first-order optimization. Math. Program. **196**, 347–388 (2022)
35. Hijazi, H., Bonami, P., Cornuéjols, G., Ouorou, A.: Mixed-integer nonlinear programs featuring "on/off" constraints. Comput. Optim. Appl. **52**, 537–558 (2012)
36. Jeon, H., Linderoth, J., Miller, A.: Quadratic cone cutting surfaces for quadratic programs with on-off constraints. Discret. Optim. **24**, 32–50 (2017)
37. Kılınç-Karzan, F., Küçükyavuz, S., Lee, D.: Joint chance-constrained programs and the intersection of mixing sets through a submodularity lens. Math. Program. **195**, 283–326 (2022)
38. Lobo, M.S., Vandenberghe, L., Boyd, S., Lebret, H.: Applications of second-order cone programming. Linear Algebra Appl. **284**, 193–228 (1998)
39. Mahajan, A., Leyffer, S., Linderoth, J., Luedtke, J., Munson, T.: Minotaur: a mixed-integer nonlinear optimization toolkit. ANL/MCS-P8010-0817, Argonne National Lab (2017)
40. Manzour, H., Küçükyavuz, S., Shojaie, A.: Integer programming for learning directed acyclic graphs from continuous data. INFORMS J. Optim. **3**, 46–73 (2020)
41. Nemhauser, G.L., Wolsey, L.A.: Integer and Combinatorial Optimization. Wiley, Hoboken (1988)
42. Nemhauser, G.L., Wolsey, L.A., Fisher, M.L.: An analysis of approximations for maximizing submodular set functions—I. Math. Program. **14**, 265–294 (1978)
43. Nguyen, T.T., Richard, J.P.P., Tawarmalani, M.: Deriving convex hulls through lifting and projection. Math. Program. **169**(2), 377–415 (2018)
44. Orlin, J.B.: A faster strongly polynomial time algorithm for submodular function minimization. Math. Program. **118**, 237–251 (2009)
45. Padberg, M.W., Van Roy, T.J., Wolsey, L.A.: Valid linear inequalities for fixed charge problems. Oper. Res. **33**(4), 842–861 (1985)
46. Pochet, Y.: Valid inequalities and separation for capacitated economic lot sizing. Oper. Res. Lett. **7**, 109–115 (1988)
47. Richard, J.P.P., Tawarmalani, M.: Lifting inequalities: a framework for generating strong cuts for nonlinear programs. Math. Program. **121**, 61–104 (2010)
48. Shi, X., Prokopyev, O.A., Zeng, B.: Sequence independent lifting for the set of submodular maximization problem. In: International Conference on Integer Programming and Combinatorial Optimization, pp. 378–390. Springer (2020)
49. Tjandraatmadja, C., Anderson, R., Huchette, J., Ma, W., Patel, K.K., Vielma, J.P.: The convex relaxation barrier, revisited: tightened single-neuron relaxations for neural network verification. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M.F., Lin, H. (eds.) Advances in Neural Information Processing Systems, vol. 33, p. 21,675-21,686. Curran Associates Inc., Red Hook (2020)
50. Van Roy, T.J., Wolsey, L.A.: Valid inequalities for mixed 0–1 programs. Discret. Appl. Math. **14**, 199–213 (1986)

51. Wei, L., Gómez, A., Küçükyavuz, S.: On the convexification of constrained quadratic optimization problems with indicator variables. In: International Conference on Integer Programming and Combinatorial Optimization, pp. 433–447. Springer (2020)
52. Wei, L., Gómez, A., Küçükyavuz, S.: Ideal formulations for constrained convex optimization problems with indicator variables. Math. Program. **192**, 57–88 (2022)
53. Wei, L., Atamtürk, A., Gómez, A., Küçükyavuz, S.: On the convex hull of convex quadratic optimization problems with indicators (2022). arXiv:2201.00387
54. Wolsey, L.A.: Submodularity and valid inequalities in capacitated fixed charge networks. Oper. Res. Lett. **8**, 119–124 (1989)
55. Wu, B., Sun, X., Li, D., Zheng, X.: Quadratic convex reformulations for semicontinuous quadratic programming. SIAM J. Optim. **27**, 1531–1553 (2017)
56. Wu, H.H., Küçükyavuz, S.: Maximizing influence in social networks: a two-stage stochastic programming approach that exploits submodularity (2015). arXiv:1512.04180
57. Xie, W., Deng, X.: Scalable algorithms for the sparse ridge regression. SIAM J. Optim. **30**, 3359–3386 (2020)
58. Yu, J., Ahmed, S.: Maximizing a class of submodular utility functions with constraints. Math. Program. **162**(1–2), 145–164 (2017)
59. Yu, J., Ahmed, S.: Polyhedral results for a class of cardinality constrained submodular minimization problems. Discret. Optim. **24**, 87–102 (2017)
60. Yu, Q., Küçükyavuz, S.: A polyhedral approach to bisubmodular function minimization. Oper. Res. Lett. **49**, 5–10 (2021)
61. Zheng, X., Sun, X., Li, D.: Improving the performance of MIQP solvers for quadratic programs with cardinality and minimum threshold constraints: a semidefinite program approach. INFORMS J. Comput. **26**, 690–703 (2014)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.