Mathematical Programming
https://doi.org/10.1007/s10107-021-01734-y

FULL LENGTH PAPER

Series B

Ideal formulations for constrained convex optimization problems with indicator variables

Linchuan Wei[1] · **Andrés Gómez**[2] · **Simge Küçükyavuz**[1]

Received: 29 June 2020 / Accepted: 24 October 2021

Abstract

Motivated by modern regression applications, in this paper, we study the convexification of a class of convex optimization problems with indicator variables and combinatorial constraints on the indicators. Unlike most of the previous work on convexification of sparse regression problems, we simultaneously consider the nonlinear non-separable objective, indicator variables, and combinatorial constraints. Specifically, we give the convex hull description of the epigraph of the composition of a one-dimensional convex function and an affine function under arbitrary combinatorial constraints. As special cases of this result, we derive ideal convexifications for problems with hierarchy, multi-collinearity, and sparsity constraints. Moreover, we also give a short proof that for a separable objective function, the perspective reformulation is ideal independent from the constraints of the problem. Our computational experiments with sparse regression problems demonstrate the potential of the proposed approach in improving the relaxation quality without significant computational overhead.

Keywords Convexification · Perspective formulation · Indicator variables · Combinatorial constraints

Mathematics Subject Classification 90C11 (Mixed-integer optimization) · 49M27 (decomposition methods) · 90C25 (convex optimization)

This research is supported, in part, by ONR grant N00014-19-1-2321, and NSF grants 1818700, 2006762, and 2007814.
A preliminary version of this work appeared in [60].

✉ Andrés Gómez
gomezand@usc.edu

Linchuan Wei
LinchuanWei2022@u.northwestern.edu

Simge Küçükyavuz
simge@northwestern.edu

[1] Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL, USA

[2] Daniel J. Epstein Department of Industrial and Systems Engineering, University of Southern California, Los Angeles, CA, USA

Published online: 02 December 2021

1 Introduction

Given a set $Q \subseteq \{0,1\}^p$, a vector $h \in \mathbb{R}^p$ such that $h_i \neq 0$, for all $i \in [p] := \{1, \ldots, p\}$, and a convex function $f : \mathbb{R} \to \mathbb{R}$, we study the set

$$Z_Q = \left\{(z, \beta, t) \in Q \times \mathbb{R}^p \times \mathbb{R} \mid f\left(h^\top \beta\right) \leq t, \beta_i(1 - z_i) = 0, \forall i \in [p]\right\}.$$

In set Z_Q above, z is a vector of indicator variables with $z_i = 1$ if $\beta_i \neq 0$, and the set Q encodes combinatorial constraints on the indicator variables. We assume without loss of generality that $f(0) = 0$, since this assumption can always be satisfied after subtracting the constant term $f(0)$.

The motivation to study Z_Q stems from sparse regression problem: Given a set of observations $(x_i, y_i)_{i=1}^n$ where $x_i \in \mathbb{R}^p$ are the features corresponding to observation i and $y_i \in \mathbb{R}$ is its associated response variable, inference with a sparse linear model can be modeled as the optimization problem

$$\min_{z,\beta} \sum_{i=1}^{n} f\left(y_i, x_i^\top \beta\right) + \lambda \rho(\beta) \tag{1a}$$

$$\text{s.t. } \beta_i(1 - z_i) = 0, \qquad i \in [p] \tag{1b}$$

$$\beta \in \mathbb{R}^p, \ z \in Q \subseteq \{0,1\}^p, \tag{1c}$$

where β is a vector of regression coefficients, f is a loss function, $\lambda \geq 0$ is a regularization parameter and ρ is regularization function. Often, $f(\beta) = \left(y_i - x_i^\top \beta\right)^2$, in which case (1) is referred to as sparse least squares regression, and typical choices of ρ include ℓ_0, ℓ_1, or ℓ_2 regularizations.

If Q is defined via a q-sparsity constraint, $Q = \left\{z \in \{0,1\}^p \mid \sum_{i=1}^p z_i \leq q\right\}$, then problem (1) reduces to the best subset selection problem [48], a fundamental problem in statistics. Nonetheless, constraints other than the cardinality constraint arise in several statistical problems. Bertsimas and King [10] suggest imposing constraints of the form $\sum_{i \in S} z_i \leq 1$ for some $S \subseteq [p]$ to prevent multicollinearity; Carrizosa et al. [18] use similar constraints to capture nested categorical variables. Constraints of the form $z_i \leq z_j$ can be used to impose strong hierarchy relationships, and constraints of the form $z_i \leq \sum_{j \in H \subseteq [p]} z_j$ can be used for weak hierarchy relationships [14]. In group variable selection, indicator variables of regression coefficients of variables in the same group are linked, see [43]. Manzour et al. [47] and Küçükyavuz et al. [46] impose that the indicator variables, which correspond to edges in an underlying graph, do not define cycles—a necessary constraint for inference problems with causal graphs. Cozad et al. [21] suggest imposing a variety of constraints in both the continuous and discrete variables to enforce priors from human experts.

Problem (1) is $\mathcal{NP}$-hard even for a q-sparsity constraint [50], and is often approximated with a convex surrogate such as lasso [39,55]. Solutions with better statistical properties than lasso can be obtained from non-convex continuous approximations [29,63]. Alternatively, it is possible to solve (1) to optimality via branch-and-bound methods [11,20]. In all cases, most of the approaches for (1) have focused on the q-

sparsity constraint (or its Lagrangian relaxation). For example, a standard technique to improve the relaxations of (1) revolves around the use of the *perspective reformulation* [1,19,26,27,30–32,34,36,42,54,61,64], an ideal formulation of a separable quadratic function with indicators (but no additional constraints). Recent work on obtaining ideal formulations for non-separable quadratic functions [4–6,27,35,44] also ignores additional constraints in Q.

There is a recent research thrust on studying constrained versions of (1). Dong et al. [25] study problem (1) from a continuous optimization perspective (after projecting out the discrete variables), see also [24]. Hazimeh and Mazumder [40] give specialized algorithms for the natural convex relaxation of (1) where Q is defined via strong hierarchy constraints. Several results exist concerning the convexification of nonlinear optimization problems with constraints [3,8,15–17,45,49,52,56–59], but such methods in general do not deliver ideal, compact or closed-form formulations for the specific case of problem (1) with structured feasible regions. In a recent work closely related to the setting considered here, Xie and Deng [62] prove that the perspective formulation is *ideal* if the objective is quadratic and separable, and Q is defined by a q-sparsity constraint. In a similar vein, Bacci et al. [7] show that the perspective reformulations for convex differentiable functions are tight for 1-sum compositions, and they use this result to show that they are ideal under unit commitment constraints. However, similar results for more general (non-separable) objective functions or constraints are currently not known.

Our contributions. In this paper, we provide a first study (from a convexification perspective) of the interplay between *non-separable* convex objectives and combinatorial constraints on the indicator variables. Specifically, we derive the convex hull description of Z_Q: the result is stated in terms of the convexification of the combinatorial set Q, but places no assumptions on its form. Using this result, we develop ideal formulations for settings in which the logical constraints on the indicator variables encode sparsity constraints or the so-called strong and weak hierarchy relations. In addition, we generalize the result in [62] and [7] to arbitrary constraints on z for *separable* convex functions f, in our setting. We show the computational benefit of the proposed approach on constrained regression problems with hierarchical relations.

An earlier version of this work appeared in [60], where we only considered separable and rank-one convex *quadratic* functions, and sparsity and strong hierarchy constraints. Furthermore, in [60], our proofs of the convexification results use the structure of each of the sets considered, whereas in the present paper, we give a unifying technique that generalizes to any combinatorial set for functions that are not necessarily quadratic. Finally, here, we expand on our preliminary computational experiments in [60] with additional datasets, conduct a further analysis on the choices of the regularization parameters, and perform computations with sparse logistic regression.

Notation. Given a one-dimensional convex function $f : \mathbb{R} \to \mathbb{R}$, we adopt the convention that $0f(\beta/0) = \lim_{z\to 0^+} zf(\beta/z)$. Using this convention, the function $zf(\beta/z)$ for $z \geq 0$ is the closure of the perspective function of f, and is convex. Let $\mathbf{0}$ and $\mathbf{1}$ be vectors of conformable dimension with all zeros and ones, respectively, and let e_i denote the ith unit vector of appropriate dimension with 1 in the ith component and zeros elsewhere. For a set Q, we denote by $\text{conv}(Q)$ its convex hull and by $\text{cl conv}(Q)$

the closure of its convex hull. Given two vectors u, v of same dimensions, we let $u \circ v$ denote the Hadamard product of u and v, i.e., $(u \circ v)_i = u_i v_i$.

2 Convexification of Z_Q

Observe that in set Z_Q, the coefficients of β can be scaled and negated if necessary to ensure $h_i = 1$ for all $i \in [p]$. Therefore, in the derivation of ideal formulations in this section, we assume, without loss of generality, that

$$Z_Q = \left\{(z, \beta, t) \in Q \times \mathbb{R}^p \times \mathbb{R} \mid f\left(\mathbf{1}^\top \beta\right) \leq t, \beta_i(1 - z_i) = 0, \forall i \in [p]\right\}.$$

We also assume, without loss of generality, that for every $i \in [p]$ there exists $z \in Q$ such that $z_i = 1$, as otherwise $z_i = \beta_i = 0$ can be fixed and the corresponding variables can be removed.

For a given set Q, let $Q^0 = Q \backslash \{\mathbf{0}\}$ or, equivalently, $Q^0 = \{z \in Q \mid \sum_{i=1}^p z_i \geq 1\}$. As we show in the subsequent discussion, the convexification of the set Z_Q relies on the characterization of $\text{conv}(Q^0)$. To this end, we first establish such a characterization.

Proposition 1 *The convex hull of Q^0 admits a description as*

$$\text{conv}(Q^0) = \text{conv}(Q) \bigcap \{z \mid \pi^\top z \geq 1, \ \forall \pi \in \mathcal{F}\}, \tag{2}$$

where $\mathcal{F}$ is a finite subset of $\mathbb{R}^p$.

Proof Let $\pi^\top z \geq \pi_0$ be an arbitrary valid inequality for $\text{conv}(Q^0)$. If $\pi_0 > 0$, then $\frac{1}{\pi_0}\pi^\top z \geq 1$ is an equivalent inequality satisfying the conditions in (2). Otherwise, if $\pi_0 \leq 0$, then the inequality does not cut off $\mathbf{0}$ and is thus valid for Q and $\text{conv}(Q)$. Therefore, it follows that $\text{conv}(Q) \subseteq \{z \mid \pi^\top z \geq \pi_0\}$, and inequality $\pi^\top z \geq \pi_0$ is either already a facet of $\text{conv}(Q)$, or is implied by the facets $\text{conv}(Q)$. Finally, finiteness of $\mathcal{F}$ follows since $\text{conv}(Q^0)$ is a polyhedron. □

Note that if $\mathbf{0} \notin Q$, then $\mathcal{F} = \emptyset$. In practice, a set $\mathcal{F}$ of minimal cardinality is preferred. Since $\text{conv}(Q)$ and $\text{conv}(Q_0)$ may have an exponential number of facets, set $\mathcal{F}$ may be exponentially large as well. In such cases, inequalities from $\mathcal{F}$ can be generated if violated in an iterative fashion, as is standard in a cutting plane algorithm. Note that even if $\text{conv}(Q)$ is simple, $\text{conv}(Q_0)$ may contain an exponential number of facets. Nonetheless, in such cases, $\text{conv}(Q_0)$ admits a compact extended formulation [2], which in turn implies that separation of the inequalities in $\mathcal{F}$ can be done in polynomial time.

Intuitively, one may think of $\mathcal{F}$ as the set of "new" facets of $\text{conv}(Q^0)$ that are not facets of $\text{conv}(Q)$. If $\text{conv}(Q)$ and $\text{conv}(Q^0)$ have the same dimension, this intuition is correct. However, if the dimension of $\text{conv}(Q^0)$ is less than the dimension of $\text{conv}(Q)$, it may be the case that $\text{conv}(Q^0) \subseteq \{z : \pi^\top z = 1\}$ for some $\pi \in \mathcal{F}$, and thus this inequality is not a facet. For example, if $Q = \{0, 1\}$, then $\text{conv}(Q) = [0, 1]$, $Q^0 = \text{conv}(Q^0) = \{1\}$ and $\mathcal{F} = \{1\}$, but the inequality $z \geq 1$ is not a facet of the 0-dimensional polyhedron $\text{conv}(Q^0)$.

The description of cl conv(Z_Q) depends on the structure of Q, and is critically dependent on whether the variables can be partitioned into multiple mutually exclusive components. We formalize this characteristic next.

Definition 1 For $i, j \in [p]$, $i \neq j$, define $i \sim j$ if there exists some $z \in Q$ such that $z_i = z_j = 1$. Define the graph $G_Q = (V, E)$ where $V = [p]$ and $\{i, j\} \in E$ if and only if $i \sim j$.

2.1 The connected case

In this section, we provide ideal formulations in the original space of variables when graph G_Q in Definition 1 is connected. This assumption is satisfied in most of the practical applications we consider, see §3. Later, in §2.2, we build upon the results of this section to derive ideal formulations when G_Q is not necessarily connected.

Before we propose a class of valid inequalities for Z_Q, we give a lemma.

Lemma 1 *For a one-dimensional proper convex function $f : \mathbb{R} \to \mathbb{R}$ with effective domain* $\text{dom}(f) = \mathbb{R}$, $f(0) = 0$ *and its perspective $g(x, t) = tf(\frac{x}{t}) : \mathbb{R}^2 \to \mathbb{R}$, if $0 < t_1 \leq t_2$, then $g(x, t_1) \geq g(x, t_2)$ for all $x \in \mathbb{R}$.*

Proof It suffices to show that the function $\phi(x) = g(x, t_1) - g(x, t_2)$ is non-decreasing in $[0, +\infty]$ and non-increasing in $[-\infty, 0]$. Since $\text{dom}(f) = \mathbb{R}$, f is continuous over $\mathbb{R}$ so is $\phi(x)$. Also, by convexity, we know that the right-derivative of $f(x)$ exists and is non-decreasing. Thus, $\phi'_+(x) = f'_+(\frac{x}{t_1}) - f'_+(\frac{x}{t_2}) \geq 0$ for all $x \in [0, +\infty]$. A continuous function with non-negative right-derivative is non-decreasing [38]. For $x \in [-\infty, 0]$, the left-derivative of ϕ is $\phi'_-(x) = f'_-(\frac{x}{t_1}) - f'_-(\frac{x}{t_2}) \leq 0$, and similarly, $\phi(x)$ is non-increasing in $[-\infty, 0]$. □

Proposition 2 *The inequalities*

$$t \geq (\pi^\top z) f\left(\frac{\boldsymbol{1}^\top \beta}{\pi^\top z}\right), \quad \forall \pi \in \mathcal{F} \tag{3}$$

are valid for Z_Q for any finite set $\mathcal{F} \subseteq \mathbb{R}^p$ satisfying (2).

Proof First, observe that if $\mathbf{0} \notin Q$, then $\mathcal{F} = \emptyset$ and the statement is superfluous. Suppose, $\mathcal{F} \neq \emptyset$. We consider two cases. If $z \neq \mathbf{0}$, then we have $\pi^\top z \geq 1$ for $\pi \in \mathcal{F}$. Then, from Lemma 1, $(\pi^\top z) f\left(\frac{\mathbf{1}^\top \beta}{\pi^\top z}\right) \leq f\left(\mathbf{1}^\top \beta\right) \leq t$. Hence the inequality is valid. Finally, if $z = \mathbf{0}$, then $\beta = \mathbf{0}$ in Z_Q. Therefore,

$$t \geq f\left(\mathbf{1}^\top \beta\right) = f(0) = 0 = \lim_{\zeta \to 0^+} \zeta f(0/\zeta) = (\pi^\top z) f\left(\frac{\mathbf{1}^\top \beta}{\pi^\top z}\right),$$

and the inequality is valid. □

We now describe the closure of the convex hull of Z_Q under the assumption that graph G_Q described in Definition 1 is connected.

Theorem 1 *If the graph G_Q given in Definition 1 is connected, then*

$$\text{cl conv}(Z_Q) = \Big\{(z,\beta,t) \in [0,1]^p \times \mathbb{R}^p \times \mathbb{R} \mid z \in \text{conv}(Q),\ t \geq f(\boldsymbol{I}^\top \beta), \\ t \geq (\pi^\top z) f\left(\frac{\boldsymbol{I}^\top \beta}{\pi^\top z}\right),\ \ \forall \pi \in \mathcal{F}\Big\} \tag{4}$$

for any finite set $\mathcal{F} \subseteq \mathbb{R}^p$ satisfying (2).

Note that if $\mathbf{0} \notin Q$, i.e., $\mathcal{F} = \emptyset$, then Theorem 1 states that the description of cl conv(Z_Q) is obtained simply by dropping the complementarity constraints $\beta_i(1 - z_i) = 0, \forall i \in [p]$ and independently taking the convex hull of Q. Otherwise, since the description of cl conv(Z_Q) requires a new inequality for every element of $\mathcal{F}$, a minimal description of $\mathcal{F}$ is certainly preferred from a computational standpoint. If conv(Q^0) is full-dimensional, the strongest nonlinear inequalities (3) are obtained from facets of conv(Q^0). Moreover, in many situations, it may not be possible to have a full description of conv(Q) or conv(Q^0); nonetheless, in those cases, it may be possible to obtain a facet $\bar{\pi}^\top x \geq 1$ of conv(Q^0), and Theorem 1 ascertains that the valid inequality

$$t \geq (\bar{\pi}^\top z) f\left(\frac{\mathbf{1}^\top \beta}{\bar{\pi}^\top z}\right) \tag{5}$$

is not dominated by any other inequality of a similar form, and that inequalities of this form are sufficient to describe cl conv(Z_Q). In Appendix 1 we focus on the special case where conv(Q) admits a compact representation but conv(Q_0) has exponentially many inequalities: We show how to use a compact extended formulation of conv(Q_0) to derive the description of cl conv(Z_Q) in a higher dimensional space.

Before proving Theorem 1, we give a lemma used in the proof.

Lemma 2 *$z \in \text{conv}(Q)$ if and only if there exists some $\alpha \in [0,1]$ and $z^0 \in \text{conv}(Q^0)$ such that $z = \alpha z^0$.*

Proof Note that if $\mathbf{0} \notin Q$, then the result holds trivially by letting $\alpha = 1$. Therefore, we will assume that $\mathbf{0} \in Q$.

($\Rightarrow$) Let $z \in \text{conv}(Q)$. So we can write z as a convex combination of the extreme points of Q. Specifically, we distinguish between the feasible points $z^i \in Q^0$ for $i \in \mathcal{I}$ and the origin. In particular, there exists $\gamma \geq \mathbf{0}$ with $\sum_{i \in \mathcal{I} \cup \{0\}} \gamma_i = 1$, such that

$$z = \gamma_0 \mathbf{0} + \sum_{i \in \mathcal{I}} \gamma_i z^i = \left(\sum_{i \in \mathcal{I}} \gamma_i\right) \sum_{i \in \mathcal{I}} \frac{\gamma_i}{\sum_{i \in \mathcal{I}} \gamma_i} z^i.$$

Letting $\alpha = \sum_{i \in \mathcal{I}} \gamma_i$, the result follows.

($\Leftarrow$) Let $z = \alpha z^0$ for some $\alpha \in [0,1]$ and $z^0 \in \text{conv}(Q^0)$; by definition, we can expand z^0 as $z^0 = \sum_{i \in \mathcal{I}} \gamma_i z^i$, a convex combination of $z^i \in Q^0$. By adding the term $(1 - \alpha)\mathbf{0}$, we have $z = (1 - \alpha)\mathbf{0} + \sum_{i \in \mathcal{I}} \alpha \gamma_i z^i$. □

We are now ready to prove Theorem 1.

Proof of Theorem 1 Define Y as the set described by (4). Let $a, b \in \mathbb{R}^p$, $c \in \mathbb{R}$, and consider the two optimization problems

$$\min_{z,\beta,t} a^\top z + b^\top \beta + ct \quad \text{subject to} \quad (z, \beta, t) \in Z_Q, \text{ and} \tag{6}$$

$$\min_{z,\beta,t} a^\top z + b^\top \beta + ct \quad \text{subject to} \quad (z, \beta, t) \in Y. \tag{7}$$

We show that there exists a solution (z, β, t) optimal for both problems, and that the corresponding objective values of both problems coincide.

- *Simple cases* :

If $c < 0$, then both (6) and (7) are unbounded. To see this, let $z = \beta = \mathbf{0}$, and $t = \kappa$, where $\kappa \geq 0$. This solution is feasible for both (6) and (7). Letting $\kappa \to \infty$, the objective goes to minus infinity.

If $c = 0$ and $b \neq \mathbf{0}$, then let $z_j = 1$ for some $j \in [p]$ such that $b_j \neq 0$, and let β_j go to plus or minus infinity depending on whether b_j is negative or positive, respectively, while keeping $\beta_i = 0$ for $i \neq j$. Again, the objective goes to minus infinity.

If $c = 0$ and $b = \mathbf{0}$, then these two problems reduce to minimizing $a^\top z$ over $\text{conv}(Q)$ and thus (6) and (7) are equivalent.

If $c > 0$, then we assume, without loss of generality, that $c = 1$ by scaling. If there exists $i_0 \neq j_0$ such that $b_{i_0} \neq b_{j_0}$, then there exist some i and j in a path from i_0 and j_0 in G_Q such that $i \sim j$ and $b_i \neq b_j$, and without loss of generality, we assume $b_i < b_j$. Furthermore, there exists some $z \in Q$ such that $z_i = z_j = 1$. Then we take such a vector z, we let β be a vector of zeros except for $\beta_i = -\beta_j = \kappa$ for some $\kappa > 0$, and we let $t = f(\mathbf{1}^\top \beta) = 0$. Such a triplet (z, β, t) is in Z_Q and Y, and by letting $\kappa \to \infty$, the objective goes to minus infinity. Therefore, we assume in the sequel that $b_i = \bar{b}$ for all $i \in [p]$.

- *Case $c = 1$ and $b = \bar{b}\mathbf{1}$* :

We now show that for $b = \bar{b}\mathbf{1}$ problem (7) either has a finite optimal solution that is in set Z_Q or is unbounded. Note that (7) is equivalent to:

$$\begin{aligned} \min_{z,\beta} \quad & a^\top z + \bar{b}\left(\mathbf{1}^\top \beta\right) + \max\left\{ f\left(\mathbf{1}^\top \beta\right), \max_{\pi \in \mathcal{F}} \left\{ \left(\pi^\top z\right) f\left(\frac{\mathbf{1}^\top \beta}{\pi^\top z}\right) \right\} \right\} \\ \text{s.t.} \quad & z \in \text{conv}(Q), \end{aligned}$$

and, from Lemma 1, it further simplifies to

$$\min_{z,\beta} \quad a^\top z + \bar{b}\left(\mathbf{1}^\top \beta\right) + \min_{\pi \in \mathcal{F}} \{\pi^\top z, 1\} f\left(\frac{\mathbf{1}^\top \beta}{\min_{\pi \in \mathcal{F}} \{\pi^\top z, 1\}}\right) \tag{8a}$$

$$\text{s.t.} \quad z \in \text{conv}(Q). \tag{8b}$$

Let $f^* : \mathbb{R} \to \mathbb{R}$ be the convex conjugate of function f, i.e., $f^*(\gamma) = \sup_{x \in \mathbb{R}} \gamma x - f(x)$, and let $\Gamma = \{\gamma \in \mathbb{R} : f^*(\gamma) < \infty\}$ be the domain of f^*. Note that if $-\bar{b} \notin \Gamma$,

it follows that both (6) and (7) are unbounded. Thus, we assume in the sequel that $-\bar{b} \in \Gamma$.

Observe that, given $w > 0$, the convex conjugate of the function $wf(x/w)$ is $wf^*(\gamma)$. Hence, from Fenchel inequality, we find that, for any β, z such that $\pi^\top z > 0$, and $\gamma \in \Gamma$,

$$\min_{\pi \in \mathcal{F}} \{\pi^\top z, 1\} f\left(\frac{\mathbf{1}^\top \beta}{\min_{\pi \in \mathcal{F}} \{\pi^\top z, 1\}}\right) \geq \gamma(\mathbf{1}^\top \beta) - \min_{\pi \in \mathcal{F}} \{\pi^\top z, 1\} f^*(\gamma). \tag{9}$$

Furthermore, for $\pi^\top z = 0$ for some $\pi \in \mathcal{F}$, if the left hand side of (9) is infinity, then the inequality holds trivially; otherwise, if the left hand side of (9) is $0 f((\mathbf{1}^\top \beta)/0) = \lim_{z \to 0^+} z f((\mathbf{1}^\top \beta)/z) = d$ with $|d| < \infty$, then by continuity of the functions at both sides of the inequality, (9) is satisfied.

Using (9) with $\gamma = -\bar{b}$ to lower bound the last term in (8a), we obtain the relaxation

$$\begin{aligned} \min_{z,\beta,t} \quad & a^\top z + \bar{b}\left(\mathbf{1}^\top \beta\right) + \left(-\bar{b}(\mathbf{1}^\top \beta) - \min_{\pi \in \mathcal{F}} \{\pi^\top z, 1\} f^*(-\bar{b})\right) \\ \text{s.t.} \quad & z \in \text{conv}(Q), \end{aligned}$$

or, equivalently,

$$\min_{z} \quad a^\top z + \max_{\pi \in \mathcal{F}} \{1 - \pi^\top z, 0\} f^*(-\bar{b}) - f^*(-\bar{b}) \tag{10a}$$

$$\text{s.t.} \quad z \in \text{conv}(Q). \tag{10b}$$

We will first prove that relaxation (10) admits an optimal solution integral in z, and then we will show that the lower bound from the relaxation is in fact tight.

Note that if $\mathbf{0} \notin Q$, then $\mathcal{F} = \emptyset$ and there exists an optimal integer solution $z^* \in Q$ to the relaxation (10) with objective value $a^\top z^* - f^*(-\bar{b})$.

Now consider the case that $\mathbf{0} \in Q$. Let z^* be an optimal solution of (10), and consider two subcases.

- *Subcase* (i) : First, suppose that $1 - \pi^\top z^* \leq 0$ for all $\pi \in \mathcal{F}$. In this case, (10) is equivalent to

$$\min_{z} \quad a^\top z - f^*(-\bar{b}) \tag{11a}$$

$$\text{s.t.} \quad \pi^\top z \geq 1 \qquad \forall \pi \in \mathcal{F} \tag{11b}$$

$$z \in \text{conv}(Q). \tag{11c}$$

From Proposition 1, the feasible region of (11) is precisely $\text{conv}(Q^0)$, thus problem (11) admits an optimal integer solution $z^* \in Q^0$ with objective value $a^\top z^* - f^*(-\bar{b})$.

- *Subcase* (ii) : Let $\bar{\pi} \in \arg\min_{\pi \in \mathcal{F}} \pi^\top z^*$, and suppose that $1 - \bar{\pi}^\top z^* > 0$. In this case, problem (10) is equivalent to

$$\ell = \min_{z} \quad a^\top z - (\bar{\pi}^\top z) f^*(-\bar{b}) \tag{12a}$$

$$\text{s.t.} \quad \pi^\top z \geq \bar{\pi}^\top z \qquad \forall \pi \in \mathcal{F} \tag{12b}$$

$$z \in \text{conv}(Q). \tag{12c}$$

Note that $f^*(-\bar{b}) = \sup_{x \in \mathbb{R}} -\bar{b}x - f(x) \geq 0$, because $x = 0$ is a possible solution to the supremum problem and $f(0) = 0$. Since $z = \mathbf{0}$ is feasible for (12), we find that the objective value $\ell \leq 0$. If $\ell = 0$, then $z^* = \mathbf{0}$ is optimal and the proof is complete. Suppose now that $\ell < 0$. Observe from Lemma 2 that $z^* = \alpha z_0$ for some $z_0 \in \text{conv}(Q^0)$ and $\alpha \in (0, 1)$—the case $\alpha = 1$ is excluded, since $1 - \bar{\pi}^\top z_0 \leq 0$ for any $z_0 \in \text{conv}(Q^0)$. Consequently, the point $\bar{z} = z_0 = \frac{z^*}{\alpha}$, with objective value $\bar{\ell} = \ell/\alpha < \ell$ is feasible for (12) with better objective value than z^*, resulting in a contradiction.

From subcases (i) and (ii), we see that either $z^* = \mathbf{0}$ is feasible and optimal for relaxation (10) (with objective value 0), or that there exists an optimal integer solution z^* with objective value $a^\top z^* - f^*(-\bar{b})$, regardless of whether $\mathbf{0} \in Q$ or not. We now prove that the lower bound provided by the relaxation (10) is tight, by finding $\beta^* \in \mathbb{R}^p$ such that $(z^*, \beta^*, f(\mathbf{1}^\top \beta^*))$ is feasible for (6) with the same objective value as (10). If $z^* = \mathbf{0}$, then clearly $(\mathbf{0}, \mathbf{0}, 0)$ is optimal for (6) with objective value 0, and we now focus on the case $z^* \neq \mathbf{0}$. Let $\bar{x} \in \arg\sup_{x \in \mathbb{R}} -\bar{b}x - f(x)$ and suppose that $\bar{x}$ exists, i.e., sup can be changed to max, and observe that $f^*(-\bar{b}) = -\bar{b}\bar{x} - f(\bar{x})$, or in other words

$$a^\top z - f^*(-\bar{b}) = a^\top z + \bar{b}\bar{x} + f(\bar{x}).$$

Since $z^* \neq \mathbf{0}$, there exists i such that $z^*_i = 1$. Setting $\beta^*_i = \bar{x}$, $\beta^*_j = 0$ for $j \neq i$, we find that the point $(z^*, \beta^*, f(\beta^*))$ is feasible for both (6) and (7), and since its objective value is the same as the lower bound obtained from (10), it is optimal for both problems. Now suppose that $\bar{x}$ above does not exist, but $(\bar{x}^1, \bar{x}^2, \ldots)$ is a sequence of points such that $-\bar{b}\bar{x}^i - f(\bar{x}^i) \to f^*(-\bar{b})$. In this case, using identical arguments as above, we find a sequence of feasible points with objective value converging to $a^\top z^* - f^*(-\bar{b})$: thus, the latter corresponds to the infimum of (7) and the relaxation is tight. □

2.2 The general case

In this section, we give ideal formulations for Z_Q when graph G_Q in Definition 1 has several connected components. Given the graph $G_Q = (V, E)$, let $V_1, V_2, \ldots, V_k$ be the vertex partition of connected components of G_Q. Let β_{V_ℓ} represent the subvector of β corresponding to indices V_ℓ. Then

$$\forall (z, \beta, t) \in Z_Q,\ f(\mathbf{1}^\top \beta) = \sum_{\ell=1}^{k} f(\mathbf{1}^\top \beta_{V_\ell}),$$

because we cannot have two indices i, j from different connected components such that $z_i = z_j = 1$. In other words, if $\beta_i \neq 0$ for some $i \in V_\ell$, $\ell \in [k]$, then $\beta_j = 0$ for all $j \in [p] \backslash V_\ell$.

For any $\ell = 1, \ldots, k$, define the projection of the binary set Q onto V_ℓ as

$$Q_\ell = \{z \in \{0,1\}^p \mid z \in Q, z_i = 0, \ \forall i \notin V_\ell\},$$

let $Q_\ell^0 = Q_\ell \backslash \{\mathbf{0}\}$ and note that, using arguments identical to those of Proposition 1, each $\text{conv}(Q_\ell^0)$ admits a description as

$$\text{conv}(Q_\ell^0) = \text{conv}(Q_\ell) \bigcap \{z \mid \pi^\top z \geq 1, \ \forall \pi \in \mathcal{F}_\ell\}$$

for some finite sets $\mathcal{F}_\ell \subseteq \mathbb{R}^p$. Note that $k > 1$ and $\mathbf{0} \notin Q_\ell$ for some $\ell \in [k]$ implies that for all $z \in Q$, $z_i = 0$ whenever $i \notin V_\ell$. Therefore, we assume that $\mathbf{0} \in Q_\ell$ and $\mathcal{F}_\ell \neq \emptyset$ for all $\ell \in [k]$. Furthermore, note that $\text{conv}(Q_\ell)$ can be described as a system of linear inequalities, i.e., $A^\ell z^\ell \leq \delta^\ell$ for all $\ell \in [k]$.

We now give the main result of this section, namely a tight extended formulation for cl conv(Z_Q) when G_Q has several connected components.

Theorem 2

$$\begin{aligned}\text{cl conv}(Z_Q) = \text{proj}_{(z,\beta,t)} \Big\{ & (z, \beta, \hat{z}, \hat{\beta}, \alpha, \hat{t}, t) \in [0,1]^p \times \mathbb{R}^p \times [0,1]^{pk} \times \mathbb{R}^{pk} \times \mathbb{R}^k_+ \times \mathbb{R}^k \times \mathbb{R} \mid \\ & \sum_{\ell=1}^k \alpha_\ell = 1, t = \sum_{\ell=1}^k \hat{t}^\ell, z = \sum_{\ell=1}^k \hat{z}^\ell, \beta = \sum_{\ell=1}^k \hat{\beta}^\ell, A^\ell \hat{z}^\ell \leq \delta^\ell \alpha_\ell, \ \forall \ell \in [k], \\ & \hat{t}^\ell \geq \alpha_\ell f\left(\frac{\boldsymbol{1}^\top \hat{\beta}^\ell}{\alpha_\ell}\right), \ \hat{t}^\ell \geq (\pi^\top \hat{z}^\ell) f\left(\frac{\boldsymbol{1}^\top \hat{\beta}^\ell}{\pi^\top \hat{z}^\ell}\right), \ \forall \pi \in \mathcal{F}_\ell, \ \forall \ell \in [k] \Big\}.\end{aligned}$$

Proof Observe that $Z_Q = \bigcup_{\ell=1}^k Z_{Q_\ell}$ and by Theorem 1, $(\hat{t}^\ell, \hat{\beta}^\ell, \hat{z}^\ell) \in \text{cl conv}(Z_{Q_\ell})$ if and only if

$$\begin{aligned} & f(\mathbf{1}^\top \hat{\beta}^\ell) - \hat{t}^\ell \leq 0, \\ & (\pi^\top \hat{z}^\ell) f\left(\frac{\mathbf{1}^\top \hat{\beta}^\ell}{\pi^\top \hat{z}^\ell}\right) - \hat{t}^\ell \leq 0, \forall \pi \in \mathcal{F}_\ell \\ & \hat{z}^\ell \in \text{conv}(Q_\ell). \end{aligned}$$

Now we see that cl conv(Z_{Q_ℓ}) has a representation in the form

$$\text{cl conv}(Z_{Q_\ell}) = \{(\hat{t}^\ell, \hat{\beta}^\ell, \hat{z}^\ell) \mid G^\ell(\hat{t}^\ell, \hat{\beta}^\ell, \hat{z}^\ell) \leq 0\},$$

where each component function of G^ℓ is closed and convex. Then using Theorem 1 in [19], we obtain a description of cl conv(Z_Q) in a higher-dimensional space by taking the perspective of G^ℓ:

$$z = \sum_{\ell=1}^{k} \hat{z}^{\ell} \tag{13a}$$

$$\beta = \sum_{\ell=1}^{k} \hat{\beta}^{\ell} \tag{13b}$$

$$t = \sum_{\ell=1}^{k} \hat{t}^{\ell} \tag{13c}$$

$$1 = \sum_{\ell=1}^{k} \alpha_{\ell} \tag{13d}$$

$$\hat{t}^{\ell} \geq \alpha_{\ell} f\left(\frac{\mathbf{1}^{\top}\hat{\beta}^{\ell}}{\alpha_{\ell}}\right) \qquad \forall \ell \in [k] \tag{13e}$$

$$\hat{t}^{\ell} \geq (\pi^{\top}\hat{z}^{\ell}) f\left(\frac{\mathbf{1}^{\top}\hat{\beta}^{\ell}}{\pi^{\top}\hat{z}^{\ell}}\right) \qquad \forall \ell \in [k], \pi \in \mathcal{F}_{\ell} \tag{13f}$$

$$A^{\ell}\hat{z}^{\ell} \leq \delta^{\ell}\alpha_{\ell} \qquad \forall \ell \in [k]. \tag{13g}$$

Hence, the result follows. □

3 Special cases

In this section, we use Theorems 1 and 2 to derive ideal formulations for Z_Q under various constraints defining Q. Direct proofs of Propositions 4, 5 and 6 were given in the preliminary version of this paper [60] for the special case of convex quadratic functions.

3.1 Unconstrained case

Consider the unconstrained case where $Q_{\mathrm{u}} = \{0, 1\}^p$ and

$$Z_{Q_{\mathrm{u}}} = \left\{(z, \beta, t) \in \{0, 1\}^p \times \mathbb{R}^{p+1} \mid f(h^{\top}\beta) \leq t, \beta_i(1 - z_i) = 0, \ \forall i \in [p]\right\}.$$

Proposition 3

$$\text{cl conv}(Z_{Q_u}) = \left\{(z, \beta, t) \in [0, 1]^p \times \mathbb{R}^{p+1} \mid f(h^{\top}\beta) \leq t, \ (\boldsymbol{1}^{\top}z) f\left(\frac{h^{\top}\beta}{\boldsymbol{1}^{\top}z}\right) \leq t\right\}.$$

Proof In this case set $Q_{\mathrm{u}}^0 = \{0, 1\}^p \backslash \{\mathbf{0}\}$ and $\text{conv}(Q_{\mathrm{u}}^0) = \{z \in [0, 1]^p \mid \mathbf{1}^{\top}z \geq 1\}$. Thus $\mathcal{F} = \{\mathbf{1}\}$ in Theorem 1, corresponding to the valid inequality $\mathbf{1}^{\top}z \geq 1$ defining $\text{conv}(Q_{\mathrm{u}}^0)$, and the result follows. □

Note that Proposition 3 generalizes existing results in the literature: if $p = 1$ and function f is one-dimensional, then Proposition 3 reduces to the perspective reformulation [19]; if $p \geq 2$ and f is quadratic, then Proposition 3 reduces to the rank-one strengthening derived in [5].

3.2 Cardinality constraint

Consider sets defined by the cardinality constraint,

$$Q_{\mathrm{c}} = \left\{ z \in \{0,1\}^p \mid \mathbf{1}^\top z \leq q \right\}.$$

Clearly, $\mathrm{conv}(Q_{\mathrm{c}}) = \left\{ z \in [0,1]^p \mid \mathbf{1}^\top z \leq q \right\}$ for any positive integer q. We now prove that, under mild conditions, ideal formulations are achieved by strengthening only the nonlinear objective.

Proposition 4 *If $q \geq 2$ and integer, then*

$$\mathrm{cl\,conv}(Z_{Q_c}) = \left\{ (z,\beta,t) \in [0,1]^p \times \mathbb{R}^{p+1} \mid \boldsymbol{1}^\top z \leq q,\ \ f(h^\top \beta) \leq t, \right. \\ \left. \left(\boldsymbol{1}^\top z\right) f\left(\frac{h^\top \beta}{\boldsymbol{1}^\top z}\right) \leq t \right\}.$$

Proof Note that if $q \geq 2$, then $G_{Q_{\mathrm{c}}}$ is a complete graph, hence $i \sim j$ for all $i, j \in [p], i \neq j$. Furthermore, $\mathrm{conv}(Q_{\mathrm{c}}^0) = \{z \in [0,1]^p : 1 \leq \mathbf{1}^\top z \leq q\}$. Hence $\mathcal{F} = \{\mathbf{1}\}$. Then the result follows from Theorem 1. □

The assumption that $q \geq 2$ in Proposition 4 is necessary. As we show next, if $q = 1$, then it is possible to strengthen the formulation with a valid inequality that uses the information from the cardinality constraint, which was not possible for $q > 1$. Note that the case $q = 1$ is also of practical interest, as set Q_{c} with $q = 1$ arises for example when preventing multi-collinearity [10] or when handling nested categorical variables [18].

Proposition 5 *If $q = 1$, then*

$$\mathrm{cl\,conv}(Z_{Q_c}) = \left\{ (z,\beta,t) \in [0,1]^p \times \mathbb{R}^{p+1} \mid \boldsymbol{1}^\top z \leq q,\ \sum_{i \in [p]} z_i f\left(\frac{h_i \beta_i}{z_i}\right) \leq t \right\}.$$

Proof First, observe that if $q = 1$, then $G_{Q_{\mathrm{c}}}$ is fully disconnected and it decomposes into p nodes, one for each variable $z_i, i \in [p]$: thus, in Theorem 2, we find that $\hat{z}_i^\ell \neq 0$ and $\hat{\beta}_i^\ell \neq 0$ if and only if $\ell = i$. In addition, because each component $\ell \in [p]$ has a single variable $\hat{z}_i^i$ for $\ell = i$, $A^\ell \hat{z}^\ell \leq \delta^\ell$ is given by $\hat{z}_i^i \leq 1$. Moreover, we find that $\mathcal{F}_i = \{1\}$ for all $i \in [p]$ in Theorem 2. Thus, from Theorem 2, we find that

$$\operatorname{cl}\operatorname{conv}(Z_{Q_c}) = \operatorname{proj}_{(z,\beta,t)}\Big\{(z,\beta,\hat{z},\hat{\beta},\alpha,\hat{t},t) \mid \sum_{i=1}^{n}\alpha_i = 1, t = \sum_{i=1}^{n}\hat{t}^i,$$
$$z_i = \hat{z}_i^i,\ \beta_i = \hat{\beta}_i^i,\ \hat{z}_i^i \le \alpha_i,\ \forall i \in [p],$$
$$\hat{t}^i \ge \alpha_i f\left(\frac{\hat{h}_i\beta_i^i}{\alpha_i}\right),\ \hat{t}^i \ge \hat{z}_i^i f\left(\frac{h_i\hat{\beta}_i^i}{\hat{z}_i^i}\right),\ \forall i \in [p]\Big\}.$$

Constraints $\hat{z}_i^i \le \alpha_i$ imply that $\hat{z}_i^i f\left(\frac{\hat{\beta}_i^i}{\hat{z}_i^i}\right) \ge \alpha_i f\left(\frac{\hat{\beta}_i^i}{\alpha_i}\right)$. Finally, variables $\hat{z}_i^i$ and $\hat{\beta}_i^i$ can be substituted with z_i and β_i, variables α_i can be projected out (resulting in the inequality $\mathbf{1}^\top z \le 1$), and the result follows. □

3.3 Strong hierarchy constraints

We now consider the hierarchy constraints. Hierarchy constraints arise from regression problems under the model (1), where the random variables include individual features as well as variables representing the interaction (usually pairwise) between a subset of these features given by a collection $\mathcal{P}$ of subsets of $[p]$. More formally, let the random variable $\theta(S)$ represent the (multiplicative) interaction of the features $i \in S$ for some subset $S \subseteq [p]$. Under this setting, the strong hierarchy constraints

$$\theta(S) \ne 0 \implies \beta_i \ne 0,\ \forall i \in S \tag{14}$$

have been shown to improve statistical performance [14,40] by ensuring that interaction terms are considered only if all corresponding features are present in the regression model. Strong hierarchy constraints can be enforced via the constraints $z(S) \le z_i$ for all $i \in S$, where $z(S) \in \{0, 1\}$ is an indicator variable such that $\theta(S)(1 - z(S)) = 0$. Thus, in order to devise strong convex relaxations of problems with hierarchy constraints, we study the set

$$Q_{\text{sh}} = \left\{z \in \{0,1\}^p \mid z_p \le z_i,\ \forall i \in [p-1]\right\}.$$

Note that in Q_{sh} we identify S with $[p-1]$, $z(S)$ with z_p and $\theta(S)$ with β_p; since p is arbitrary, this identification is without loss of generality.

To establish the convex hull of $Z_{Q_{\text{sh}}}$, we give a lemma that characterizes $\operatorname{conv}(Q_{\text{sh}}^0)$. First, observe that

$$\sum_{i\in[p-1]} z_i - (p-2)z_p \ge 1 \tag{15}$$

is a valid inequality for Q_{sh}^0. To see this, note that for $z \ne \mathbf{0}$, if $z_p = 0$, then we must have $\sum_{i\in[p-1]} z_i \ge 1$, and if $z_p = 1$, then we must have $\sum_{i\in[p-1]} z_i = p - 1$, so the validity follows.

Lemma 3

$$\text{Conv}(Q^0_{sh}) = \left\{ z \in [0,1]^p \mid \sum_{i\in[p-1]} z_i - (p-2)z_p \geq 1,\ z_p \leq z_i,\ \forall i \in [p-1] \right\}.$$

Proof Let

$$Q_g = \left\{ z \in [0,1]^p \mid \sum_{i\in[p-1]} z_i - (p-2)z_p \geq 1,\ z_p \leq z_i,\ \forall i \in [p-1] \right\}.$$

We will first show that the extreme points of Q_g are integral. Then we will prove that $\text{conv}(Q_{\text{sh}} \backslash \mathbf{0}) = Q_g$.

Suppose z^* is an extreme point of Q_g. Observe that if z^*_p is equal to 1, then $z^*_i = 1$ for all $i \in [p-1]$. If z^*_p is equal to 0, then the constraint matrix defining Q_g is totally unimodular, thus all extreme points of Q_g with $z^*_p = 0$ are integral. If constraint (15) is not tight at an extreme point, then because the remaining constraint matrix defining Q_g is totally unimodular, the corresponding extreme point of Q_g is integral. Therefore, it suffices to consider extreme points where (15) holds at equality and $0 < z^*_p < 1$.

Now suppose $\sum_{i\in[p-1]} z^*_i - (p-2)z^*_p = 1$ and $1 > z^*_p > 0$. We first show that $z^*_i = 1$ for at most one coordinate $i \in [p-1]$. If $z^*_i = z^*_j = 1$ for $i \neq j$, then

$$\sum_{\ell\in[p-1]} z^*_\ell - (p-2)z^*_p = z^*_i + \sum_{\ell\in[p-1],\ell\neq i} (z^*_\ell - z^*_p) \geq z^*_i + (z^*_j - z^*_p) > z^*_i = 1, \tag{16}$$

where the first inequality follows from dropping terms $z^*_\ell - z^*_p \geq 0$ with $\ell \neq j$, and the second inequality follows from the assumption $z^*_j = 1$ and $z^*_p < 1$. Since (16) contradicts $\sum_{i\in[p-1]} z^*_i - (p-2)z^*_p = 1$, it follows that $z^*_i = 1$ for at most one coordinate $i \in [p-1]$.

Next, observe that if $z^*_i = z^*_p$ for all $i \in [p-1]$, then $\sum_{i\in[p-1]} z^*_i - (p-2)z^*_p = z^*_p < 1$. Therefore, the largest element in $z^*_i, i \in [p-1]$ has to be strictly greater than z^*_p. Finally, we now show that we can perturb z^*_p and the $p-2$ smallest elements in $z^*_i, i \in [p-1]$ by a small quantity ϵ and remain in Q_g. The equality $\sum_{i\in[p-1]} z_i - (p-2)z_p = 1$ clearly holds after the perturbation. And, adding a small quantity ϵ to z^*_p and the $p-2$ smallest elements in $z^*_i, i \in [p-1]$ does not violate the hierarchy constraint since the largest element in $z^*_i, i \in [p-1]$ is strictly greater than z^*_p. Finally, since $z^*_i \geq z^*_p > 0, \forall i \in [p-1]$, subtracting a small quantity ϵ does not violate the non-negativity constraint. Thus, we can write z^* as a convex combination of two points in Q_g, which is a contradiction.

To see that $Q_g = \text{conv}(Q^0_{\text{sh}})$, first, observe that $\mathbf{0} \notin Q_g$. Also, (15) is a valid inequality for Q^0_{sh}. Furthermore, we just showed that the extreme points of Q_g are integral, hence $Q_g = \text{conv}(Q^0_{\text{sh}})$. □

Now we are ready to give an ideal formulation for $Z_{Q_{\text{sh}}}$.

Proposition 6 *The closure of the convex hull of $Z_{Q_{sh}}$ is given by*

$$\text{cl conv}(Z_{Q_{sh}}) = \Big\{(z,\beta,t) \in [0,1]^p \times \mathbb{R}^{p+1} \mid f(h^\top \beta) \le t,\ z_p \le z_i, \forall i \in [p-1], \left(\sum_{i\in[p-1]} z_i - (p-2)z_p\right) f\left(\frac{h^\top \beta}{\sum_{i\in[p-1]} z_i - (p-2)z_p}\right) \le t\Big\}.$$

Proof First, observe that the constraint matrix defining Q_{sh} is totally unimodular, so $\text{conv}(Q_{\text{sh}}) = \left\{z \in [0,1]^p \mid z_p \le z_i,\ \forall i \in [p-1]\right\}$. Note that $G_{Q_{\text{sh}}}$ is a complete graph, hence $i \sim j$ for all $i, j \in [p], i \ne j$. Hence, from Lemma 3, $\mathcal{F} = \{(1,\ldots,1,-(p-2))\}$. Then the result follows from Theorem 1. □

3.4 Weak hierarchy

Consider the strong hierarchy relation (14), which requires all variables in the set S to have non-zero coefficients to capture a multiplicative effect, $\theta(S)$ on the response variable y. The weak hierarchy relation [14] is a relaxation of the strong hierarchy relation to address the interaction between random variables in the same subset S by requiring

$$\theta(S) \ne 0 \implies \beta_i \ne 0, \quad \text{for } \textit{some } i \in S.$$

Using similar arguments as before, we formulate the weak hierarchy relation as $z_p \le \sum_{i\in[p-1]} z_i$, in other words, $z_1, z_2, \ldots, z_{p-1} = 0 \implies z_p = 0$. The corresponding constrained indicator variable set is thus defined by

$$Q_{\text{wh}} = \left\{z \in \{0,1\}^p \mid z_p \le \sum_{i\in[p-1]} z_i\right\}.$$

Note that $\mathbf{1} \in Q_{\text{wh}}$, thus the graph $G_{Q_{\text{wh}}}$ is connected and Theorem 1 can be used to derive the convex hull.

Proposition 7

$$\text{cl conv}(Z_{Q_{wh}}) = \Big\{(z,\beta,t) \in [0,1]^p \times \mathbb{R}^{p+1} \mid f(h^\top \beta) \le t, z_p \le \sum_{i\in[p-1]} z_i, \left(\sum_{i\in[p-1]} z_i\right) f\left(\frac{h^\top \beta}{\sum_{i\in[p-1]} z_i}\right) \le t\Big\}.$$

Proof First, observe that the constraint matrix defining Q_{wh} is totally unimodular, hence $\text{conv}(Q_{\text{wh}}) = \left\{z \in [0,1]^p \mid z_p \le \sum_{i\in[p-1]} z_i\right\}$. Clearly, $\sum_{i\in[p-1]} z_i \ge 1$ is valid for Q^0_{wh} since $z_1 = \cdots = z_{p-1} = 0 \implies z_p = 0$. It suffices to show that

$$\operatorname{conv}(Q^0_{\text{wh}}) = \left\{ z \in [0,1]^p \mid \sum_{i \in [p-1]} z_i \geq 1 \right\}. \tag{17}$$

All extreme points of the polyhedron on the right-hand side of (17) are integral, because the associated constraint matrix is an interval matrix with integral right-hand side. The result follows from Theorem 1. □

4 A note on separable functions

In this section, we demonstrate that the proof technique used in §2 can be extended to separable functions with constraints, resulting in relatively simple proofs generalizing existing results in the literature.

Given a partition of $[p] = \bigcup_{j=1}^{\ell} V_j$ and convex functions $f_j : \mathbb{R}^{V_j} \to \mathbb{R}$ such that $f_j(\mathbf{0}) = 0$, consider the epigraph of a separable function of the form:

$$W = \Big\{ z \in Q \subseteq \{0,1\}^\ell, \beta \in \mathbb{R}^p, t \in \mathbb{R} \mid \sum_{j=1}^{\ell} f_j(\beta_{V_j}) \leq t, \\ \beta_i(1 - z_j) = 0, \ \forall j \in [\ell], i \in V_j \Big\}.$$

As Theorem 3 below shows, ideal formulations of W can be obtained by applying the perspective reformulation on the separable nonlinear terms and, *independently*, strengthening the continuous relaxation of Q. Let

$$Y_s = \left\{ (z, \beta, t) \in \mathbb{R}^{\ell+p+1} \mid \sum_{j=1}^{\ell} z_j f_j\left(\frac{\beta_{V_j}}{z_j}\right) \leq t, \ z \in \operatorname{conv}(Q) \right\}.$$

Theorem 3 *Y_s is the closure of the convex hull of W:* $\operatorname{cl}\operatorname{conv}(W) = Y_s$.

Proof Validity of the corresponding inequality in Y_s follows directly from the validity of the perspective reformulation. For any $(a, b, c) \in \mathbb{R}^{\ell+p+1}$ consider the following two problems

$$\min \ \ a^\top z + b^\top \beta + ct \qquad \text{subject to } (z, \beta, t) \in W, \tag{18}$$

and

$$\min \ \ a^\top z + b^\top \beta + ct \qquad \text{subject to } (z, \beta, t) \in Y_s. \tag{19}$$

It suffices to show that (18) and (19) are equivalent, i.e., there exists an optimal solution of (19) that is optimal for (18) with the same objective value. As before, we may assume that $c = 1$ without loss of generality. For $j \in [\ell]$, let $f_j^* : \mathbb{R}^{V_j} \to \mathbb{R}$ be the convex conjugate of function f_j, i.e.,

$$f_j^*(\gamma) = \sup_{\beta \in \mathbb{R}^{V_j}} \gamma^\top \beta - f_j(\beta),$$

and let $\Gamma_j = \left\{\gamma \in \mathbb{R}^{V_j} : f_j^*(\gamma) < \infty\right\}$. From Fenchel's inequality corresponding to the perspective function, we find that for any $\beta \in \mathbb{R}^{V_j}$, $z_j \geq 0$ and $\gamma \in \Gamma_j$,

$$z_j f_j\left(\frac{\beta}{z_j}\right) \geq \gamma^\top \beta - z_j f_j^*(\gamma). \tag{20}$$

Observe that both (18) and (19) are unbounded if $-b_{V_j} \notin \Gamma_j$ for some $j \in [\ell]$. Otherwise, if $-b_{V_j} \in \Gamma_j$ for all $j \in [\ell]$, we use (20) with $\gamma = -b_{V_j}$ for each $j \in [\ell]$ to lower bound the objective of (19), resulting in the relaxation

$$\min \quad \sum_{j=1}^{\ell} \left(a_j - f_j^*(-b_{V_j})\right) z_j \tag{21a}$$

$$\text{s.t.} \quad z \in \text{conv}(Q), \tag{21b}$$

which admits an optimal solution $z^* \in Q$. Letting $\beta^*_{V_j} \in \arg\sup_{\beta \in \mathbb{R}^{V_j}} - b_{V_j}^\top \beta - f_j(\beta_{V_j})$ whenever $z_j^* = 1$ and $\beta^*_{V_j} = \mathbf{0}$ otherwise, we find a feasible solution for (18) with the same objective value. □

Theorem 3 generalizes the result of Xie and Deng [62] for $Q = \left\{z \in \{0,1\}^p \mid \sum_{i=1}^p z_i \leq q\right\}$, $V_j = \{j\}$, and $f_j(\beta_j) = \beta_j^2$ for $j \in [p]$. Theorem 3 also generalizes the result of Bacci et al. [7] for the case that f_j is convex, differentiable and certain constraint qualification conditions hold, applied to our setting. However, Bacci et al. [7] consider more general settings where multiple polyhedra are connected by a single binary variable, and under linear constraints on the continuous variables.

5 Quadratic case: implementation via semidefinite optimization

In this section we review how to implement the convexifications derived in §2 for the special case of quadratic optimization. Given observations $(x_i, y_i)_{i=1}^n$ with $x_i \in \mathbb{R}^p$ and $y_i \in \mathbb{R}$, let $X_{n \times p}$ defined as $X_{ij} = (x_i)_j$ be the model matrix, and consider least square regression problems

$$\min_{z,\beta} \ \|y - X\beta\|_2^2 + \lambda \|\beta\|_2^2 + \mu \|z\|_1 \tag{22a}$$

$$\text{s.t. } \beta_i(1 - z_i) = 0 \qquad \forall i \in [p] \tag{22b}$$

$$\beta \in \mathbb{R}^p, \ z \in Q \subseteq \{0,1\}^p, \tag{22c}$$

where the regularization terms $\lambda\|\beta\|_2^2$ and $\mu\|z\|_1$ penalize the ℓ_2-norm and ℓ_0-norm of β, respectively. A natural convexification of (22) based on the ℓ_2-regularization term $\lambda\|\beta\|_2^2$ is to directly use the perspective relaxation [13,62]

$$\min_{z,\beta} \|y - X\beta\|_2^2 + \lambda \sum_{i=1}^{p} t_i + \mu \|z\|_1 \tag{23a}$$

$$\text{s.t. } \beta_i^2 \le t_i z_i \qquad \forall i \in [p] \tag{23b}$$

$$\beta \in \mathbb{R}^p,\ z \in \text{conv}(Q). \tag{23c}$$

Formulation (23) can either be directly implemented with conic quadratic solvers [1], implemented via cutting plane methods [33] or via tailored methods specific to linear regression [41]. However, (23) is weak if λ is small.

In this paper we focus on relaxations that do not assume the presence of the ℓ_2-regularization term (but require solving an SDP). In particular, letting $B \approx \beta\beta^\top$, Dong et al. [26] propose the semidefinite relaxation of (22) given by

$$\min_{z,\beta,B} \|y\|_2^2 - 2y^\top X\beta + \langle X^\top X + \lambda I, B\rangle + \mu \sum_{i=1}^{p} z_i \tag{24a}$$

$$\text{s.t. } \begin{pmatrix} z_i & \beta_i \\ \beta_i & B_{i,i} \end{pmatrix} \succeq 0 \qquad \forall i \in [p] \tag{24b}$$

$$\begin{pmatrix} 1 & \beta^\top \\ \beta & B \end{pmatrix} \succeq 0 \tag{24c}$$

$$\beta \in \mathbb{R}^p,\ B \in \mathbb{R}^{p\times p},\ z \in \text{conv}(Q), \tag{24d}$$

which dominates the perspective relaxation (23), as well as any perspective relaxation obtained from extracting a diagonal matrix from $X^\top X + \lambda I$, e.g., using the method in [32]. We now discuss how (24) can be further strengthened.

Given any $T \subseteq [p]$, let β_T, z_T and B_T the subvectors of β and z and submatrix of B induced by T, respectively. Moreover, let Q_T be the projection of Q onto the subspace of variables in T. First, observe that in order to apply our theoretical developments to this setting, we need to extract a convex function of the form $f(h^\top \beta_T)$ for some $h \in \mathbb{R}^{|T|}$. In particular, we consider quadratic f. Note that for any h, from Theorem 1, we can obtain valid inequalities of the form

$$t \ge \frac{(h^\top \beta_T)^2}{\pi^\top z_T}, \quad \forall \pi \in \mathcal{F}_T \tag{25}$$

for some set $\mathcal{F}_T \subseteq \mathbb{R}^{|T|}$ describing Q_T^0. Inequalities (25) can then be included in formulation (24) by using the methodology given in [37], as discussed next.

For any $h \in \mathbb{R}^{|T|}$, we find that for $z \in Q_T$ and $B_T = \beta_T \beta_T^\top$ satisfying (22b),

$$\langle hh^\top, B_T\rangle = (h^\top \beta)^2 \ge \frac{(h^\top \beta_T)^2}{\pi^\top z_T}. \tag{26}$$

Observe that inequality (26) is valid for any vector h. Therefore, by optimizing over h to find the strongest inequality, we obtain

$$0 \geq \max_{h \in \mathbb{R}^{|T|}} \left\{ \frac{(h^\top \beta_T)^2}{\pi^\top z_T} - \langle hh^\top, B_T \rangle \right\}. \tag{27}$$

Inequality (27) is satisfied if and only if $h^\top \left(\beta_T \beta_T^\top / \pi^\top z_T - B_T \right) h \leq 0$ for all $h \in \mathbb{R}^{|T|}$, or, equivalently, if $B_T - \beta_T \beta_T^\top / \pi^\top z_T \succeq 0$. Using Schur complement, we conclude that constraint (27) is equivalent to

$$\begin{pmatrix} \pi^\top z_T & \beta_T^\top \\ \beta_T & B_T \end{pmatrix} \succeq 0. \tag{28}$$

Observe that inequalities (24b) are in fact special cases of (28) with $T = \{i\}, i \in [p]$.

6 Numerical results

In this section, we provide numerical results to compare relaxations of regression problems. Specifically, in §6.1 we present computations with sparse least squares regression problems with all pairwise (second-order) interactions and strong hierarchy constraints [40]; in §6.2 we present computations with logistic regression. The conic optimization problems are solved with MOSEK 8.1 solver on a laptop with a 2.0 GHz intel(R)Core(TM)i7-8550H CPU with 16 GB main memory.

6.1 Least squares regression with hierarchy constraints

In this section we focus on least squares regression problems with hierarchy constraints. A usual approach to compute estimators to statistical inference problems is either to use the relaxation of a suitable convex relaxation directly, or to round the solution obtained from such convex relaxations, see for example [5,6,9,12,26,51,62]. Thus, as a proxy to evaluate the quality of the estimators obtained, we focus on the optimality gap provided by such approaches. In §6.1.1 we discuss the relaxations used, and in §6.1.2 we discuss a simple rounding heuristic, which guarantees that the produced solutions satisfy the hierarchy constraints.

6.1.1 Formulations

Given observations $(x_\ell, y_\ell)_{\ell=1}^n$, we consider relaxations of the problem

$$\min_{z,\beta} \sum_{\ell=1}^{n} \left(y_\ell - \sum_{i=1}^{p} x_{\ell i} \beta_i - \sum_{i=1}^{p} \sum_{j=i}^{p} x_{\ell i} x_{\ell j} \beta_{ij} \right)^2 + \lambda \|\beta\|_2^2 + \mu \|z\|_1 \tag{29a}$$

$$\text{s.t. } \beta_i (1 - z_i) = 0 \quad \forall i \in [p] \tag{29b}$$

$$\beta_{ij} (1 - z_{ij}) = 0 \quad \forall i, j \in [p],\ i \leq j \tag{29c}$$

$$z_{ii} \leq z_i \quad \forall i \in [p] \tag{29d}$$

$$z_{ij} \leq z_i,\ z_{ij} \leq z_j \quad \forall i, j \in [p],\ i \leq j \tag{29e}$$

$$\beta \in \mathbb{R}^{p(p+3)/2},\ z \in \{0,1\}^{p(p+3)/2}. \tag{29f}$$

We standardize the data so that all columns have 0 mean and norm 1, i.e., $\|y\|_2 = 1$, $\|X_i\|_2^2 = 1$ for all $i \in [p]$, and $\|X_i \circ X_j\|_2^2 = 1$ for all $i \leq j$ (where $X_i \in \mathbb{R}^n$ and $(X_i)_\ell = x_{\ell i}$). Note that constraints (29d)-(29e) are totally unimodular, hence conv(Q) in (24d) can be obtained simply by relaxing integrality constraints to $0 \leq z \leq 1$.

In addition to the **optimal perspective** reformulation (24), we consider the following strengthenings.

Rank1 Inequalities (28) for all sets T of cardinality 2 using the "unconstrained" convexification given in Proposition 3. This formulation was originally proposed in [5]. The resulting semidefinite constraints are of the form

$$\begin{pmatrix} z_i + z_j & \beta_i & \beta_j \\ \beta_i & B_{i,i} & B_{i,j} \\ \beta_j & B_{i,j} & B_{j,j} \end{pmatrix} \succeq 0, \quad \begin{pmatrix} z_i + z_{jk} & \beta_i & \beta_{jk} \\ \beta_i & B_{i,i} & B_{i,jk} \\ \beta_{jk} & B_{i,jk} & B_{jk,jk} \end{pmatrix} \succeq 0,$$
$$\text{or} \quad \begin{pmatrix} z_{i_1 i_2} + z_{j_1 j_2} & \beta_{i_1 i_2} & \beta_{j_1 j_2} \\ \beta_{i_1 i_2} & B_{i_1 i_2, i_1 i_2} & B_{i_1 i_2, j_1 j_2} \\ \beta_{j_1 j_2} & B_{i_1 i_2, j_1 j_2} & B_{j_1 j_2, j_1 j_2} \end{pmatrix} \succeq 0.$$

Hier Inequalities (28) for all sets T linked by hierarchy constraints. Specifically, from constraints (29d) we add constraints with $|T| = 2$ of the form

$$\begin{pmatrix} z_i & \beta_i & \beta_{ii} \\ \beta_i & B_{i,i} & B_{i,ii} \\ \beta_{ii} & B_{i,ii} & B_{ii,ii} \end{pmatrix} \succeq 0.$$

Moreover, from constraints (29e), linking the three variables β_i, β_j and β_{ij}, we add constraints involving pairs of variables β_i and β_{ij} of the form

$$\begin{pmatrix} z_i & \beta_i & \beta_{ij} \\ \beta_i & B_{i,i} & B_{i,ij} \\ \beta_{ij} & B_{i,ij} & B_{ij,ij} \end{pmatrix} \succeq 0.$$

Constraints involving pairs of variables β_j and β_{ij} are identical and added as well. Finally, constraints considering the three variables simultaneously are added, resulting in constraints with $|T| = 3$ of the form

$$\begin{pmatrix} z_i + z_j - z_{ij} & \beta_i & \beta_i & \beta_{ij} \\ \beta_i & B_{i,i} & B_{i,j} & B_{i,ij} \\ \beta_j & B_{i,j} & B_{j,j} & B_{j,ij} \\ \beta_{ij} & B_{i,ij} & B_{j,ij} & B_{ij,ij} \end{pmatrix} \succeq 0.$$

Rank1+hier All inequalities of both **Rank1** and **Hier**.

6.1.2 Upper bounds and gaps

Given the solution of the convex relaxation, we use a simple rounding heuristic to recover a feasible solution to problem (29): we round z_i and fix it to the nearest

Table 1 Datasets

Dataset	n	p	$p(p+3)/2$
Crime	51	5	20
Diabetes	442	10	65
Wine_quality (red)	1599	11	77
Forecasting_orders	60	12	90
Housing	507	13	104
Bias_correction	7,590	18	189

integer—observe that a rounded solution always satisfies hierarchy constraints (29d)-(29e)—, and solve the resulting convex optimization problem in terms of β. Given the objective value v_ℓ of the convex relaxation and v_u of the heuristic, we can bound the optimality gap as gap $= \frac{v_u - v_\ell}{v_u} \times 100\%$.

6.1.3 Instances and parameters

We test the formulations on six datasets: Crime (from [39]), Diabetes (from [28]), Housing, Wine_quality (red), Forecasting_orders and Bias_correction (latter four from [23]). Table 1 shows the number of observations n and number of original regression variables p, as well as the total number of variables $p(p+3)/2$ after adding all second order interactions. Finally we use regularization values $(\lambda, \mu) = (0.01i, 0.01j)$ for all $0 \leq i, j \leq 30$ with $i, j \in \mathbb{Z}_+$ for all datasets but Bias_correction, for which we use for $1 \leq i \leq 20$ and $1 \leq j \leq 30$ with $i, j \in \mathbb{Z}_+$, due to its larger size and longer relaxation solution times.

6.1.4 Results

Figure 1 shows the distribution of times needed to solve the regression problems for each dataset. As expected, the optimal **perspective** formulation (24) is the fastest, as it is the simplest relaxation. We also see that formulations involving the rank-one constraints (with or without hierarchical strengthening) are more computationally demanding, taking four times longer to solve than the perspective formulation in Crime, and twice as long in the remaining five instances. In contrast, the formulation **Hier**, which includes hierarchical constraints but not the rank-one constraints, is much faster, requiring 70% more time than **perspective** in the Crime dataset, and only 10-20% more in the other instances. Indeed, there are only $\mathcal{O}(p^2)$ hierarchical constraints to be added, while there are $\mathcal{O}((p(p+3)/2)^2)$ rank-one constraints.

Figure 2 shows, for each dataset, the average optimality gaps as a function of the regularization parameter λ. Each point in the graph represents, for a given value of λ the average across all 31 values of μ. Similarly, Fig. 3 shows, for each dataset, the average optimality gap as a function of the regularization parameter μ. As expected, the optimality gaps obtained from the optimal perspective reformulation are the largest, as the relaxation (24) is dominated by all the other relaxations used. Moreover, the relaxation **rank1+hier** results in the smallest gaps, as it dominates every other relax-

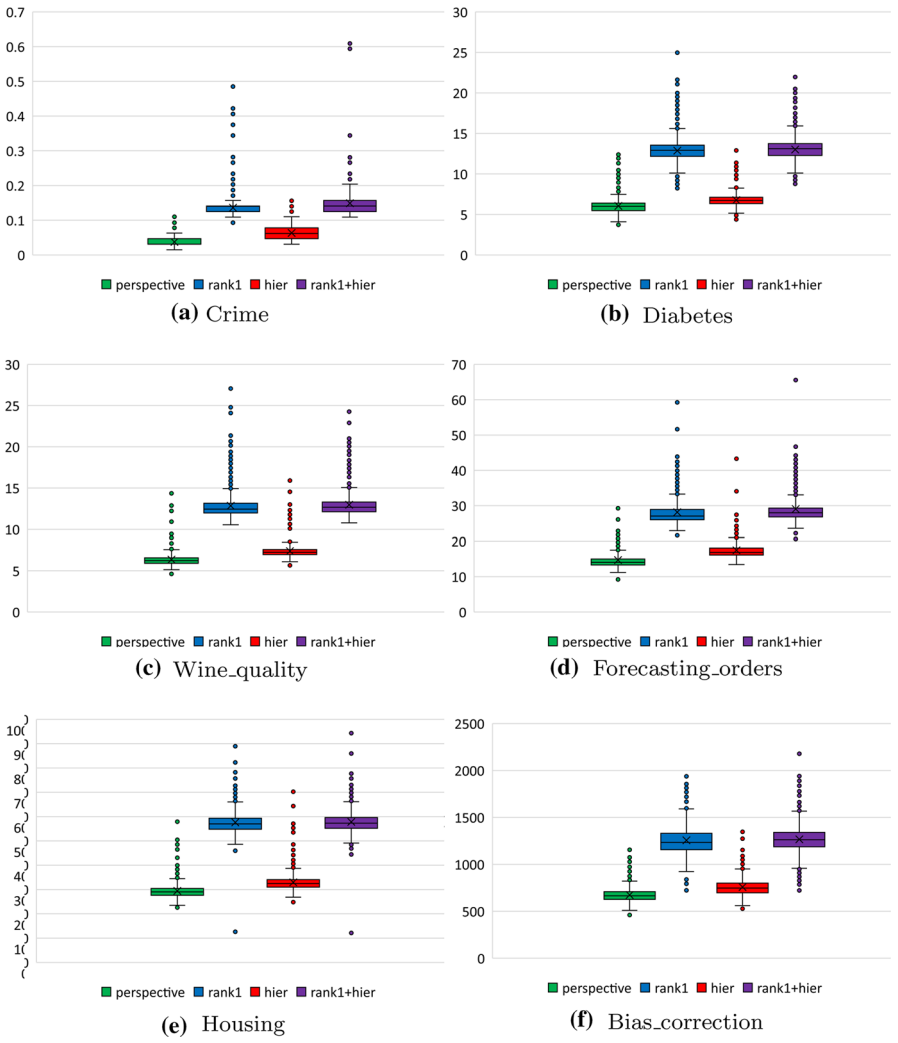

Fig. 1 Computational times in seconds

ation used. Finally, neither relaxation **rank1** nor **hier** consistently outperforms the other, although **hier** results in lower gaps overall in all datasets except Diabetes.

The relative performance of the formulations tested in terms of gap largely depends on the dataset and parameters used. In the Diabetes and Wine_quality datasets, the **perspective** reformulation is by far the worst, and all other formulations significantly improve upon it. Specifically, **rank1+hier** is slightly better than **rank1** and **hier** (which have similar strengths), but the differences are marginal—observe that **hier** achieves an almost ideal strengthening with half the computational cost of the other formulations. In contrast, in the Crime, Housing and Bias_correction datasets, **rank1** achieves only a marginal improvement over the **perspective** relaxation, while **hier** achieves a significant improvement over **rank1**, and **rank1+hier** results in a even more sub-

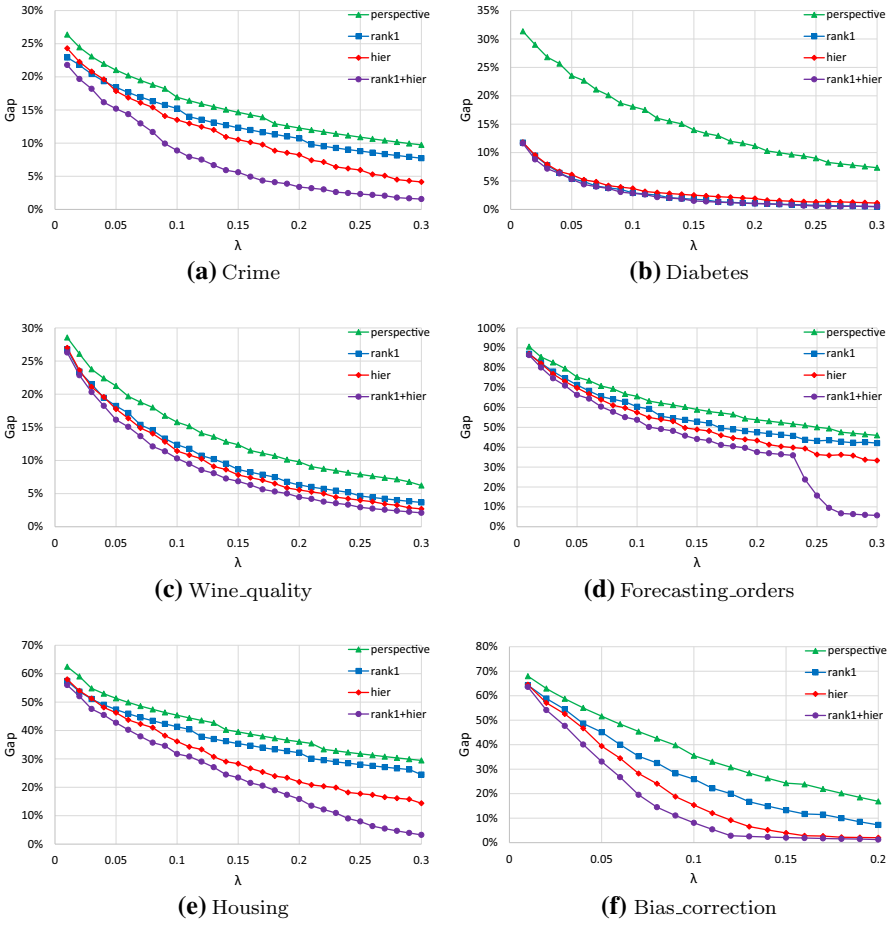

Fig. 2 Optimality gaps as a function of λ. Each point in the graph represents the average across all values of μ

stantial improvement. For example, for the Housing dataset, for $\lambda = 0.3$ the average optimality gap of **perspective** is 29%, whereas that of **rank1+hier** is 3%. Finally, in Forecasting_orders, all formulations perform similarly for $\lambda \leq 0.23$; however, for $\lambda \geq 0.24$, **rank1+hier** results in a significant improvement over the other formulations. Note that Forecasting_orders is a "fat" dataset with $n < p(p+3)/2$, which is more difficult for convexifications of the form (24) for low values of λ. Our conclusions from Fig. 3 are similar. Of particular note is the marked improvement in the optimality gaps for **rank1+hier** over other formulations (especially **perspective**) for small μ. For example, for $\mu = 0.05$, the average optimality gap of **rank1+hier** in the Crime dataset is slightly over 5%, whereas **hier** achieves over 10% gap, and **perspective** and **rank1** result in 25% gap.

Since **hier** has a similar computational cost as **perspective**, and **rank1+hier** has a virtually identical cost as **rank1**, we see that the hierarchical strengthening may lead

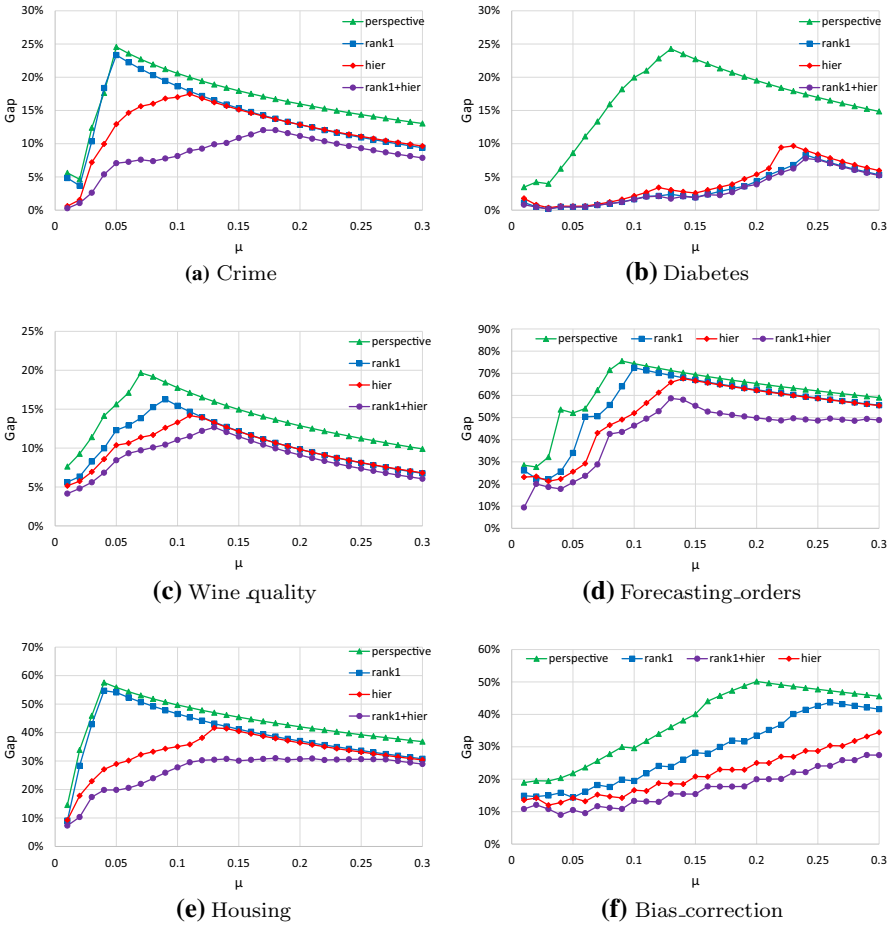

Fig. 3 Optimality gaps as a function of μ. Each point in the graph represents the average across all values of λ

to large improvements without drawbacks (whereas **rank1** requires 2–4 times more computational overhead). Indeed, the hierarchical strengthening is tailored to problem (29), while **rank1** is more general but does not exploit any structural information from the constraints.

Remark 1 (On rounding vs mixed-integer optimization) An alternative to the rounding approach used here is to use mixed-integer optimization (MIO) to solve (29) exactly. An extensive comparison between MIO and the **rank1** approach was performed in [5] in a variety of real datasets, including the Diabetes dataset used here. In summary, while MIO (using the perspective reformulation) was found to be more effective for large values of the parameter λ, simple rounding of the **rank1** relaxation was already sufficient to prove smaller optimality gaps than MIO if λ is small. We refer to reader to [5] for additional details.

6.2 Sparse logistic regression

To illustrate the convexification of non-quadratic functions derived in §2, we consider ℓ_0-regularized logistic regression problems. Specifically, given a classification problem with data $(x_i, y_i)_{i=1}^n$ where $x_i \in \mathbb{R}^p$ and $y_i \in \{-1, 1\}$, sparse logistic regression calls for solving the problem [22,53]

$$\min_{z,\beta} \quad (1-\lambda)\sum_{i=1}^{n} \log\left(1+\exp(-y_i x_i^\top \beta)\right) + \lambda \sum_{i=1}^{p} z_i \tag{30a}$$

$$\text{s.t.} \quad \beta_i(1-z_i) = 0, \qquad \forall i \in [p], \tag{30b}$$

where $0 \le \lambda \le 1$ is a regularization coefficient that controls the balance between the error and the ℓ_0-penalty. Note that the natural convex relaxation of (30), obtained by dropping the indicator variables z (or, equivalently, adding big M constraints $|\beta_i| \le M z_i$ with $M \to \infty$), is convex.

To date, there are limited results concerning convexifications of (30), especially when compared with the sparse least squares regression problem (22), due to: *(i)* non-existence of separable terms amenable to the perspective relaxation; *(ii)* lack of convexifications for non-separable non-quadratic terms with indicators; and *(iii)* non-decomposability of the objective function into simpler terms, resulting in similar convexifications as those discussed in §5. In this paper we provided the first convexifications for non-quadratic non-separable functions, addressing issue *(ii)*. In this section we illustrate that *if* the observations x_i are sufficiently sparse, then a direct application of Theorem 1 results in substantial improvements over the natural relaxation, circumventing issues *(i)* and *(iii)*.

6.2.1 Formulations

A direct application of Theorem 1, corresponding to strengthening each error term $\log\left(1+\exp(-y_i x_i^\top \beta)\right) \le t_i$ individually, yields the following "rank-one" relaxation of the sparse logistic regression problem (30):

$$\min_{z,\beta,t} \quad (1-\lambda)\left(\sum_{i=1}^{n} t_i + n\log(2)\right) + \lambda \sum_{i=1}^{p} z_i \tag{31a}$$

$$\text{s.t.} \quad t_i \ge \log(1+\exp(-y_i x_i^\top \beta)) - \log(2), \quad i \in [n] \tag{31b}$$

(log-rko)

$$t_i \ge \left(\sum_{j:x_{ij}\neq 0} z_j\right) \log\left(1+\exp\left(-\frac{y_i x_i^\top \beta}{\sum_{j:x_{ij}\neq 0} z_j}\right)\right) - \left(\sum_{j:x_{ij}\neq 0} z_j\right)\log(2), \quad i \in [n] \tag{31c}$$

$$z_i \in [0,1], \quad i \in [n]. \tag{31d}$$

We can write (31) as a conic optimization problem using the exponential cone

$$K_{exp} = \{(w_1, w_2, w_3) \mid w_1 \geq w_2 e^{w_3/w_2}, w_2 > 0\} \bigcup \{(w_1, 0, w_3) \mid w_1 \geq 0, w_3 \leq 0\},$$

i.e., the closure of the set of points satisfying $w_1 \geq w_2 e^{w_3/w_2}$ and $w_1, w_2 \geq 0$. Constraint (31b) is equivalent to $\exists u_1^i, v_1^i$ such that :

$$\begin{aligned} u_1^i + v_1^i &\leq 2, \\ (u_1^i, 1, -y_i x_i^\top \beta - t_i) &\in K_{exp}, \\ (v_1^i, 1, -t_i) &\in K_{exp}. \end{aligned}$$

Similarly, constraint (31c) is equivalent to $\exists u_2^i, v_2^i$ such that:

$$\begin{aligned} u_2^i + v_2^i &\leq 2\Big(\sum_{j:x_{ij}\neq 0} z_j\Big), \\ \Big(u_2^i, \sum_{j:x_{ij}\neq 0} z_j, -y_i x_i^\top \beta - t_i\Big) &\in K_{exp}, \\ \Big(v_2^i, \sum_{j:x_{ij}\neq 0} z_j, -t_i\Big) &\in K_{exp}. \end{aligned}$$

We refer to formulation (31) as **log-rko** in the sequel. We compare it with the natural convex relaxation (30) **log-nat**, corresponding to dropping contraints (31c) from the formulation. Observe that for relaxation **log-nat**, $z = \mathbf{0}$ in an optimal solution, thus resulting in the same objective value for all values of λ.

6.2.2 Lower bounds

We report lower bounds found from solving convex relaxations of (30). The optimal values of the relaxations considered are divided by $(1-\lambda)n\log(2)$. Thus the feasible solution of (30) obtained by setting $z = \beta = \mathbf{0}$ has objective value 1 (this solution may be optimal for large values of λ). The objective value of (30) also have a trivial lower bound of 0, attained if the data can be perfectly classified (observe that if $n < p$ and $\lambda \to 0$, this lower bound may in fact be attained).

6.2.3 Instances and parameters

For the synthetic datasets, we consider the case where both the input data and the true model are sparse. Let α be a parameter that controls the sparsity of features $(x_i)_{i=1}^n$. For each entry x_{ij} we either independently assign a value of zero with probability $1-\alpha$ or we sample from a standard normal distribution $\mathcal{N}(0, 1)$ with probability α. We generate a "true" sparse coefficient vector β^* with s uniformly sampled non-zero indices such that $\beta_i^* = 1$. The responses $y_i \in \{-1, 1\}$ are then generated independently from

Table 2 Scaled lower bounds for varying n, λ, and α. Each entry in the table represents the average across ten random instances

		$\alpha = 0.01$	$\alpha = 0.02$	$\alpha = 0.05$	$\alpha = 0.1$
$n = 50$					
	log-nat	0.430	0.136	0.008	0
	log-rko $\lambda = 0.1$	0.502	0.214	0.059	0.027
	$\lambda = 0.3$	0.703	0.434	0.203	0.103
	$\lambda = 0.5$	0.92	0.726	0.441	0.239
	$\lambda = 0.7$	1	0.965	0.798	0.524
$n = 100$					
	log-nat	0.392	0.164	0.003	0
	log-rko $\lambda = 0.1$	0.454	0.221	0.036	0.017
	$\lambda = 0.3$	0.626	0.383	0.133	0.066
	$\lambda = 0.5$	0.834	0.615	0.296	0.153
	$\lambda = 0.7$	0.985	0.893	0.588	0.340
$n = 200$					
	log-nat	0.519	0.328	0.220	0.232
	log-rko $\lambda = 0.1$	0.564	0.370	0.243	0.242
	$\lambda = 0.3$	0.685	0.483	0.308	0.272
	$\lambda = 0.5$	0.837	0.644	0.414	0.323
	$\lambda = 0.7$	0.974	0.859	0.598	0.434

a Bernoulli distribution with: $P(y_i = 1|x_i) = (1 + \exp(-x_i^\top \beta^*))^{-1}$. We use regularization values $\lambda \in \{0.1, 0.3, 0.5, 0.7\}$ and sparsity levels $\alpha \in \{0.01, 0.02, 0.05, 0.1\}$. Moreover, we set $p = 100$, $s = 1$ and test varying sample sizes $n = 50, 100, 200$.

6.2.4 Results

Table 2 shows the scaled lower bounds obtained via convex relaxations **log-nat** and **log-rko**. Each entry in the table corresponds to the average (over ten replications) lower bound obtained from a given relaxation for a particular combination of sparsity level α, λ, and number of observations n. Recall that **log-nat** results in the same objective regardless of the value of λ.

Compared with the natural relaxation of sparse logistic regression, the lower bound attained by (31) increases significantly when $n \leq p$. Moreover, as expected, larger improvements of **log-rko** over **log-nat** are obtained for larger values of λ, where sparsity plays a more prominent role in the objective value. The lower bounds of **log-rko** are at least 16% more than those of **log-nat** in all test cases, and sometimes substantially larger (e.g., in cases **log-nat** results in the trivial lower bound of 0). When the input data is very sparse, i.e., $\alpha = 0.01$ and $\lambda \geq 0.5$, **log-rko** results in lower bounds close to 1 or equal to 1, suggesting (and in some cases proving) that true optimal solution in those cases is $z = \beta = \mathbf{0}$. When $n > p$, **log-rko** still results

in better lower bounds than **log-nat**, although improvements are less pronounced in these cases.

7 Conclusions

In this paper, we propose a unifying convexification technique for the epigraphs of a class of convex functions with indicator variables constrained to certain polyhedral sets. We illustrate the utility of our approaches on constrained regression problems of recent interest. Our results generalize the existing results that consider only quadratic, separable or differentiable convex functions, and certain structural constraints such as cardinality or unit commitment. As future research, we plan to consider convexifications for more general functions.

Acknowledgements We thank the AE and two referees whose comments expanded and improved our computational study, and also led to the result in Appendix 1.

A The special case when conv(Q) is compact

In this section, we give an extended formulation of cl conv(Z_Q) based on an extended formulation of conv(Q_0). In particular, this alternative formulation is more favorable in cases when the number of facets of conv(Q) is polynomially bounded while conv(Q_0) has an exponential number of facets. We denote the facets of conv(Q) which do not contain zero by $\{F_\ell\}_{1\le\ell\le k}$, and we write each F_ℓ as $F_\ell := \{z \mid A_\ell z \le b_\ell\}$. Angulo et al. [2] prove that $\text{conv}(Q_0) = \text{conv}\left(\bigcup_{1\le\ell\le k} F_\ell\right)$, and a natural extended formulation of conv(Q_0) is as follows:

$$z = \sum_{\ell\in[k]} \hat{z}_\ell \tag{32a}$$

$$A_\ell \hat{z}_\ell \le \lambda_\ell b_\ell \qquad \ell \in [k] \tag{32b}$$

$$\sum_{\ell\in[k]} \lambda_\ell = 1,\ \lambda \ge \mathbf{0}. \tag{32c}$$

Theorem 4

$$\begin{aligned}&\text{cl conv}(Z_Q)\\ &= \text{proj}_{(z,\beta,t)}\left\{(z,\hat{z},\lambda,\beta,t) \in \mathbb{R}_+^{(k+1)p+k} \times \mathbb{R}^p \times \mathbb{R} \mid (32a) - (32b),\ \sum_{\ell\in[k]} \lambda_\ell \le 1,\right.\\ &\qquad\qquad \left. t \ge f(\boldsymbol{I}^\top\beta),\ t \ge (\boldsymbol{I}^\top\lambda) f\left(\frac{\boldsymbol{I}^\top\beta}{\boldsymbol{I}^\top\lambda}\right)\right\}.\end{aligned}$$

Proof Let

$$Z = \Big\{(z, \hat{z}, \lambda, \beta, t) \in \mathbb{R}_+^{(k+1)p+k} \times \mathbb{R}^p \times \mathbb{R} \mid (32a) - (32b),\ \sum_{\ell \in [k]} \lambda_\ell \leq 1, t \geq f(\mathbf{1}^\top \beta),$$
$$t \geq (\mathbf{1}^\top \lambda) f\left(\frac{\mathbf{1}^\top \beta}{\mathbf{1}^\top \lambda}\right)\Big\}.$$

First we show that $\text{proj}_{(z,\beta,t)}(Z) \subseteq \text{cl conv}(Z_Q)$. Given any $(z, \hat{z}, \lambda, \beta, t) \in Z$, note that constraints $z \in \text{conv}(Q)$ and $t \geq f(\mathbf{1}^\top \beta)$ defining $\text{cl conv}(Z_Q)$ are trivially satisfied. It remains to show that $t \geq (\pi^\top z) f\left(\frac{\mathbf{1}^\top \beta}{\pi^\top z}\right), \quad \forall \pi \in \mathcal{F}$. For each $\pi \in \mathcal{F}$, we have

$$\pi^\top z = \sum_{\ell \in [k]} \pi^\top \hat{z}_\ell = \sum_{\ell \in [k]} \lambda_\ell \pi^\top \left(\frac{\hat{z}_\ell}{\lambda_\ell}\right) \geq \sum_{\ell \in [k]} \lambda_\ell,$$

where the inequality follows from the fact that we must have either $\lambda_\ell = 0$ and $\hat{z}_\ell = \mathbf{0}$ or $\lambda_\ell > 0$ and $\frac{\hat{z}_\ell}{\lambda_\ell} \in F_\ell$ since each F_ℓ is a polytope contained in the half-space defined by inequality $\mathbf{1}^\top z \geq 1$. Thus, from Lemma 1, we have $t \geq (\mathbf{1}^\top \lambda) f\left(\frac{\mathbf{1}^\top \beta}{\mathbf{1}^\top \lambda}\right) \geq (\pi^\top z) f\left(\frac{\mathbf{1}^\top \beta}{\pi^\top z}\right), \quad \forall \pi \in \mathcal{F}$, hence $\text{proj}_{(z,\beta,t)}(Z) \subseteq \text{cl conv}(Z_Q)$.

Now, it remains to prove that $\text{cl conv}(Z_Q) \subseteq \text{proj}_{(z,\beta,t)}(Z)$. For any $(z, \beta, t) \in \text{cl conv}(Z_Q)$ if $z \in \text{conv}(Q_0)$, then there exist $\hat{z}_\ell$ and λ_ℓ that satisfy (32) and $\mathbf{1}^\top \lambda = 1$. Since $t \geq f(\mathbf{1}^\top \beta)$ for all $(z, \beta, t) \in \text{cl conv}(Z_Q)$, $(z, \beta, t) \in \text{proj}_{(z,\beta,t)}(Z)$. If $z \in \text{conv}(Q) \setminus \text{conv}(Q_0)$, then, from Lemma 2, we can write z as $z = \lambda_0 z_0$, $0 \leq \lambda_0 < 1$, and we may assume z_0 is on one of the facets of $\text{conv}(Q_0)$ defined by $\hat{\pi}^\top z_0 = 1$ for some $\hat{\pi} \in \mathcal{F}$. By definition, $\forall \pi \in \mathcal{F} \quad \pi^\top z_0 \geq \hat{\pi}^\top z_0 = 1$ which implies $\lambda_0 = \hat{\pi}^\top z = \min_{\pi \in \mathcal{F}} \pi^\top z$. Since $z_0 \in \text{conv}(Q_0)$, there exists $\hat{z}_\ell, \lambda_\ell$ such that $z_0 = \sum_{\ell \in [k]} \hat{z}_\ell$ and (32b)–(32c) hold. Then

$$\begin{aligned} z &= \sum_{\ell \in [k]} (\lambda_0 \hat{z}_\ell) \\ A_\ell(\lambda_0 \hat{z}_\ell) &\leq \lambda_0 \lambda_\ell b_\ell, && \ell \in [k] \\ \sum_{\ell \in [k]} \lambda_0 \lambda_\ell &\leq 1, \lambda \geq \mathbf{0}, \end{aligned}$$

and we have $\sum_{\ell \in [k]} \lambda_0 \lambda_\ell = \lambda_0 = \min_{\pi \in \mathcal{F}} \pi^\top z$. Using Lemma 1, we find that $t \geq (\pi^\top z) f\left(\frac{\mathbf{1}^\top \beta}{\pi^\top z}\right), \quad \forall \pi \in \mathcal{F}$ implies that $t \geq \left(\sum_{\ell \in [k]} \lambda_0 \lambda_\ell\right) f\left(\frac{\mathbf{1}^\top \beta}{\sum_{\ell \in [k]} \lambda_0 \lambda_\ell}\right) \geq \left(\sum_{\ell \in [k]} \lambda_\ell\right) f\left(\frac{\mathbf{1}^\top \beta}{\sum_{\ell \in [k]} \lambda_\ell}\right)$. Hence, $(z, \beta, t) \in \text{proj}_{(z,\beta,t)}(Z)$. □

References

1. Aktürk, M.S., Atamtürk, A., Gürel, S.: A strong conic quadratic reformulation for machine-job assignment with controllable processing times. Oper. Res. Lett. **37**(3), 187–191 (2009)
2. Angulo, G., Ahmed, S., Dey, S.S., Kaibel, V.: Forbidden vertices. Math. Oper. Res. **40**(2), 350–360 (2015)
3. Anstreicher, K.M.: On convex relaxations for quadratically constrained quadratic programming. Math. Program. **136**(2), 233–251 (2012)
4. Atamtürk, A., Gómez, A.: Strong formulations for quadratic optimization with M-matrices and indicator variables. Math. Program. **170**(1), 141–176 (2018)
5. Atamtürk, A., Gómez, A.: Rank-one convexification for sparse regression. Optimization Online. http://www.optimization-online.org/DB_HTML/2019/01/7050.html. (2019)
6. Atamtürk, A., Gómez, A., Han, S.: Sparse and smooth signal estimation: convexification of L0 formulations. J. Mach. Learn. Res. **3**, 1–43 (2021)
7. Bacci, T., Frangioni, A., Gentile, C., Tavlaridis-Gyparakis, K.: New MINLP formulations for the unit commitment problems with ramping constraints. Optimization Online. http://www.optimization-online.org/DB_FILE/2019/10/7426.pdf. (2019)
8. Belotti, P., Góez, J.C., Pólik, I., Ralphs, T.K., Terlaky, T.: A conic representation of the convex hull of disjunctive sets and conic cuts for integer second order cone optimization. In: Numerical Analysis and Optimization, pp. 1–35. Springer (2015)
9. Bertsimas, D., Cory-Wright, R., Pauphilet, J.: Mixed-projection conic optimization: A new paradigm for modeling rank constraints. arXiv preprint arXiv:2009.10395 (2020a)
10. Bertsimas, D., King, A.: OR Forum - An algorithmic approach to linear regression. Oper. Res. **64**(1), 2–16 (2016)
11. Bertsimas, D., King, A., Mazumder, R.: Best subset selection via a modern optimization lens. Ann. Stat. **44**(2), 813–852 (2016)
12. Bertsimas, D., Pauphilet, J., Van Parys, B., et al.: Sparse regression: scalable algorithms and empirical performance. Stat. Sci. **35**(4), 555–578 (2020b)
13. Bertsimas, D., Van Parys, B.: Sparse high-dimensional regression: Exact scalable algorithms and phase transitions. Ann. Statist. **1**, 300–323 (2020)
14. Bien, J., Taylor, J., Tibshirani, R.: A lasso for hierarchical interactions. Ann. Stat. **41**(3), 1111 (2013)
15. Bienstock, D., Michalka, A.: Cutting-planes for optimization of convex functions over nonconvex sets. SIAM J. Opt. **24**(2), 643–677 (2014)
16. Burer, S.: On the copositive representation of binary and continuous nonconvex quadratic programs. Math. Program. **120**(2), 479–495 (2009)
17. Burer, S., Kılınç-Karzan, F.: How to convexify the intersection of a second order cone and a nonconvex quadratic. Math. Program. **162**(1–2), 393–429 (2017)
18. Carrizosa, E., Mortensen, L., Morales, D.R.: On linear regression models with hierarchical categorical variables. Tech. rep. (2020)
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.: Learning surrogate models for simulation-based optimization. AIChE J. **60**(6), 2211–2227 (2014)
21. 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)
22. Dedieu, A., Hazimeh, H., Mazumder, R.: Learning sparse classifiers: Continuous and mixed integer optimization perspectives. J. Mach. Learn. Res. **15**, 1–4 (2021)
23. Dheeru, D., Karra Taniskidou, E.: UCI machine learning repository (2017)
24. Dong, H.: On integer and MPCC representability of affine sparsity. Oper. Res. Lett. **47**(3), 208–212 (2019)
25. Dong, H., Ahn, M., Pang, J.-S.: Structural properties of affine sparsity constraints. Math. Program. **176**(1–2), 95–135 (2019)
26. Dong, H., Chen, K., Linderoth, J.: Regularization vs. relaxation: A conic optimization perspective of statistical variable selection. arXiv preprint arXiv:1510.06083 (2015)
27. Dong, H., Linderoth, J.: On valid inequalities for quadratic programming with continuous variables and binary indicators. In: Goemans, M., Correa, J. (eds.) Integer Programming and Combinatorial Optimization, pp. 169–180. Springer, Berlin (2013)

28. Efron, B., Hastie, T., Johnstone, I., Tibshirani, R.: Least angle regression. Ann. Stat. **32**(2), 407–499 (2004)
29. Fan, J., Li, R.: Variable selection via nonconcave penalized likelihood and its oracle properties. J. Am. Stat. Assoc. **96**(456), 1348–1360 (2001)
30. Frangioni, A., Furini, F., Gentile, C.: Approximated perspective relaxations: a project and lift approach. Comput. Opt. Appl. **63**(3), 705–735 (2016)
31. Frangioni, A., Gentile, C.: Perspective cuts for a class of convex 0–1 mixed integer programs. Math. Program. **106**, 225–236 (2006)
32. Frangioni, A., Gentile, C.: SDP diagonalizations and perspective cuts for a class of nonseparable MIQP. Oper. Res. Lett. **35**(2), 181–185 (2007)
33. Frangioni, A., Gentile, C.: A computational comparison of reformulations of the perspective relaxation: SOCP vs. cutting planes. Oper. Res. Lett. **37**(3), 206–210 (2009)
34. Frangioni, A., Gentile, C., Grande, E., Pacifici, A.: Projected perspective reformulations with applications in design problems. Oper. Res. **59**(5), 1225–1232 (2011)
35. 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)
36. Günlük, O., Linderoth, J.: Perspective reformulations of mixed integer nonlinear programs with indicator variables. Math. Program. **124**, 183–205 (2010)
37. Han, S., Gómez, A., Atamtürk, A.: 2x2 convexifications for convex quadratic optimization with indicator variables. arXiv preprint arXiv:2004.07448 (2020)
38. Hardy, G.H.: Course of Pure Mathematics. Courier Dover (1908). (**Publications**)
39. Hastie, T., Tibshirani, R., Wainwright, M.: Statistical learning with sparsity: the lasso and generalizations. In: Monographs on Statistics and Applied Probability, vol. 143. Chapman and Hall/CRC (2015)
40. Hazimeh, H., Mazumder, R.: Learning hierarchical interactions at scale: a convex optimization approach. In: Chiappa, S. and Calandra, R., editors, Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics, volume 108 of Proceedings of Machine Learning Research, pp. 1833–1843. PMLR
41. Hazimeh, H., Mazumder, R., Saab, A.: Sparse regression at scale: branch-and-bound rooted in first-order optimization. arXiv preprint arXiv:2004.06152 (2020)
42. Hijazi, H., Bonami, P., Cornuéjols, G., Ouorou, A.: Mixed-integer nonlinear programs featuring "on/off" constraints. Comput. Opt. Appl. **52**(2), 537–558 (2012)
43. Huang, J., Breheny, P., Ma, S.: A selective review of group selection in high-dimensional models. Stat. Sci. Rev. J. Inst. Math. Stat. **27**(4),(2012)
44. Jeon, H., Linderoth, J., Miller, A.: Quadratic cone cutting surfaces for quadratic programs with on-off constraints. Dis. Opt. **24**, 32–50 (2017)
45. Kılınç-Karzan, F., Yıldız, S.: Two-term disjunctions on the second-order cone. In: International Conference on Integer Programming and Combinatorial Optimization, pp. 345–356. Springer. (2014)
46. Küçükyavuz, S., Shojaie, A., Manzour, H., Wei, L.: Consistent second-order conic integer programming for learning Bayesian networks. arXiv preprint arXiv:2005.14346 (2020)
47. Manzour, H., Küçükyavuz, S., Wu, H.-H., Shojaie, A.: Integer programming for learning directed acyclic graphs from continuous data. INFORMS J. Opt. **3**(1), 46–73 (2021)
48. Miller, A.: Subset selection in regression. Chapman and Hall/CRC (2002)
49. Modaresi, S., Kılınç, M.R., Vielma, J.P.: Intersection cuts for nonlinear integer programming: convexification techniques for structured sets. Math. Program. **155**(1–2), 575–611 (2016)
50. Natarajan, B.K.: Sparse approximate solutions to linear systems. SIAM J. Comput. **24**(2), 227–234 (1995)
51. Pilanci, P., Wainwright, M.J., El Ghaoui, L.: Sparse learning via Boolean relaxations. Math. Program. **151**, 63–87 (2015)
52. Richard, J.-P.P., Tawarmalani, M.: Lifting inequalities: a framework for generating strong cuts for nonlinear programs. Math. Program. **121**(1), 61–104 (2010)
53. Sato, T., Takano, Y., Miyashiro, R., Yoshise, A.: Feature subset selection for logistic regression via mixed integer optimization. Comput. Opt. Appl. **64**(3), 865–880 (2016)
54. Stubbs, R.A., Mehrotra, S.: A branch-and-cut method for 0–1 mixed convex programming. Math. Program. **86**(3), 515–532 (1999)
55. Tibshirani, R.: Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B Methodol., pp. 267–288 (1996)

56. Vielma, J.P.: Small and strong formulations for unions of convex sets from the Cayley embedding. Math. Program. **177**(1–2), 21–53 (2019)
57. Wang, A. L., Kılınç-Karzan, F.: The generalized trust region subproblem: solution complexity and convex hull results. Forthcoming in Math. Program. (2020a)
58. Wang, A.L., Kılınç-Karzan, F.: On convex hulls of epigraphs of QCQPs. In: Bienstock, D., Zambelli, G. (eds.) Integer Programming and Combinatorial Optimization, pp. 419–432. Cham. Springer International Publishing (2020b)
59. Wang, A. L., Kılınç-Karzan, F.: On the tightness of SDP relaxations of QCQPs. Forthcoming in Math. Program. (2021)
60. Wei, L., Gómez, A., Küçükyavuz, S.: On the convexification of constrained quadratic optimization problems with indicator variables. In: Bienstock, D., Zambelli, G. (eds.) Integer Programming and Combinatorial Optimization, pp. 433–447. Cham. Springer International Publishing (2020)
61. Wu, B., Sun, X., Li, D., Zheng, X.: Quadratic convex reformulations for semicontinuous quadratic programming. SIAM J. Opt. **27**(3), 1531–1553 (2017)
62. Xie, W., Deng, X.: Scalable algorithms for the sparse ridge regression. SIAM J. Opt. **30**(4), 3359–3386 (2020)
63. Zhang, C.-H.: Nearly unbiased variable selection under minimax concave penalty. Ann. Stat. **38**, 894–942 (2010)
64. 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**(4), 690–703 (2014)