

Tilburg University

Hidden Convexity in Partially Separable Optimization

Ben-Tal, A.; den Hertog, D.; Laurent, M.

Publication date:
2011

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):

Ben-Tal, A., den Hertog, D., & Laurent, M. (2011). *Hidden Convexity in Partially Separable Optimization*. (CentER Discussion Paper; Vol. 2011-070). Operations research.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

No. 2011-070

**HIDDEN CONVEXITY IN PARTIALLY SEPARABLE
OPTIMIZATION**

By Aharon Ben-Tal, Dick den Hertog, Monique Laurent

June 10, 2011

ISSN 0924-7815

Hidden convexity in partially separable optimization

Aharon Ben-Tal *

*Department of Industrial Engineering and Management,
Technion – Israel Institute of Technology, Haifa 32000, Israel
CentER Extramural Fellow, CentER, Tilburg University, The Netherlands*

Dick den Hertog

Tilburg University, Department of Econometrics and Operations Research, 5000 LE Tilburg, Netherlands

Monique Laurent

*Centrum Wiskunde & Informatica (CWI), Science Park 123, 1098 XG Amsterdam
Tilburg University, Department of Econometrics and Operations Research, 5000 LE Tilburg, Netherlands*

June 10, 2011

Abstract

The paper identifies classes of nonconvex optimization problems whose convex relaxations have optimal solutions which at the same time are global optimal solutions of the original nonconvex problems. Such a *hidden convexity* property was so far limited to quadratically constrained quadratic problems with one or two constraints. We extend it here to problems with some partial separable structure. Among other things, the new hidden convexity results open up the possibility to solve multi-stage robust optimization problems using certain *nonlinear* decision rules.

Keywords: convex relaxation of nonconvex problems, hidden convexity, partially separable functions, robust optimization

JEL classification: C61

1 Introduction

In this paper we reveal some classes of nonconvex optimization problems which nevertheless possess a hidden convexity property. By this it is meant that there is an explicit convex problem having an optimal solution which is at the same time a global optimal solution of the original nonconvex problem. Hidden convexity was discovered for problems with indefinite quadratic objective functions, and one or two quadratic constraints ([10],[13]). The hidden convex problems in these cases are generally Semi Definite Programs (SDP), or, in more specific cases, are Conic Quadratic Programs (CQP) ([3]). The main mathematical tools to derive these results are basic results on the range space of (indefinite) quadratic forms ([11]) and the so-called S-lemma ([12],

*This paper was written when the author was visiting Centrum Wiskunde & Informatica in Amsterdam, The Netherlands, as a CWI Distinguished Scientist.

[8]). Here we go beyond the quadratic case, and hence no reliance on such tools is possible anymore. The classes of problems we study are all special cases of the following general form

$$(NCP) \quad \min_{x \in \mathbb{R}^n} \{G_0(x) \mid G_l(x) \leq \rho_l, l \in L\},$$

where L is a finite index set, and, for $l \in \{0\} \cup L$:

$$G_l(x_1, \dots, x_n) = g_l(\beta_{l1}f_1(x_{[1]}), \dots, \beta_{lN}f_N(x_{[N]})) + h_l(\alpha_{l1}x_1, \dots, \alpha_{ln}x_n).$$

Here $x_{[1]}, x_{[2]}, \dots, x_{[N]}$ is a partition of the variables (x_1, x_2, \dots, x_n) into N mutually exclusive subvectors (this is the “block-separable” case considered in Section 2). Moreover, $g_l : \mathbb{R}^N \rightarrow \mathbb{R}$, $h_l : \mathbb{R}^n \rightarrow \mathbb{R}$, and $f_j : \mathbb{R}^{|[j]|} \rightarrow \mathbb{R}$ are convex functions. The source of nonconvexity in problem (NCP) comes from the possible negative signs of some of the coefficients β_{lk} . But even if all β_{lk} s are nonnegative $G_l(x)$ can be nonconvex, since g_l is not assumed to be componentwise monotone increasing.

The general methodology by which we try to expose hidden convexity in problem (NCP) is as follows. First we write (NCP) equivalently in terms of additional variables y_1, \dots, y_N :

$$\begin{cases} \min_{x \in \mathbb{R}^n, y \in \mathbb{R}^N} F_0(x, y) = g_0(\beta_{01}y_1, \dots, \beta_{0N}y_N) + h_0(\alpha_{01}x_1, \dots, \alpha_{0n}x_n) \\ \text{s.t.} \\ F_l(x, y) = g_l(\beta_{l1}y_1, \dots, \beta_{lN}y_N) + h_l(\alpha_{l1}x_1, \dots, \alpha_{ln}x_n) \leq \rho_l, l \in L \\ f_j(x_{[j]}) = y_j, j = 1, \dots, N. \end{cases} \quad (1)$$

We then relax the nonconvex constraints $f_j(x_{[j]}) = y_j$ by replacing them with

$$f_j(x_{[j]}) \leq y_j, j = 1, \dots, N, \quad (2)$$

thus obtaining the following *convex optimization problem*:

$$(CP) \quad \min\{F_0(x, y) \mid F_l(x, y) \leq \rho_l, l \in L; f_j(x_{[j]}) \leq y_j, j = 1, \dots, N\}.$$

Now, if we can prove that (CP) has an optimal solution which happens to satisfy all the constraints in (2) as *equalities* (we call such a solution an *E-solution*), then this solution is an optimal (global) solution of problem (1) and hence of problem (NCP) .

The main special case treated in the paper, for which we prove (by construction) the *existence* of an E-solution is when the functions g_l and h_l are linear and $|L| \leq 2$, or $|L| > 2$ but an additional property (“termwise parallel”) links the constraints in some sense. Another special case is when $|L| \geq 1$, g_l are general convex functions, and h_l are convex functions with certain restrictions on the signs of their partial derivatives. For this case we prove that *every* optimal solution of (CP) is an E-solution.

The above results are presented in Section 2 and are finally collected in Table 1. In Section 3 we show how the hidden convexity results offer the possibility to derive robust solutions of constraints in which the dependence on the uncertain parameters is nonlinear. Moreover, for multi-period linear problems we show that new opportunities to go beyond using just linear decision rules (as in [5], [2]) open up. In the last section the complexity of (CP) and its dual problem is studied. It is shown that in many cases polynomial complexity is achieved.

2 Exact convex relaxations of partially separable problems

We first start with the fully separable case ($N = n$). Given univariate functions $f_1, \dots, f_n : \mathbb{R} \rightarrow \mathbb{R}$ and vectors $\alpha_l, \beta_l \in \mathbb{R}^n$, (for $l \in \{0\} \cup L$), $\rho_l \in \mathbb{R}$ (for $l \in L$), we consider the following problem in n variables $x = (x_1, \dots, x_n)$ and with $m = |L|$ constraints:

$$\min \alpha_0^T x + \beta_0^T f(x) \quad \text{s.t.} \quad \alpha_l^T x + \beta_l^T f(x) \leq \rho_l \quad (l \in L). \quad (3)$$

Note that this problem is a special case of problem (NCP) in the Introduction (take $h_l(t_1, \dots, t_n) = g_l(t_1, \dots, t_n) = \sum_{i=1}^n t_i, l \in \{0\} \cup L$). We introduce new variables $y = (y_1, \dots, y_n)$ for $f(x) = (f_1(x_1), \dots, f_n(x_n))$, leading to the following reformulation of (3):

$$\min \alpha_0^T x + \beta_0^T y \quad \text{s.t.} \quad \alpha_l^T x + \beta_l^T y \leq \rho_l \quad (l \in L), \quad f(x) = y. \quad (4)$$

Now we relax the equation $y = f(x)$ by the (coordinate-wise) inequality $f(x) \leq y$, and obtain the following relaxation of (3) (and (4)):

$$\min \alpha_0^T x + \beta_0^T y \quad \text{s.t.} \quad \alpha_l^T x + \beta_l^T y \leq \rho_l \quad (l \in L), \quad f(x) \leq y. \quad (5)$$

Observe that, in the special case when each f_i is a convex function, (5) is a convex program, whereas the original problem (3) is not convex when not all β_i s are nonnegative.

The following definition plays a key role in our analysis.

Definition 2.1. *A feasible solution (x, y) to problem (5) is an E-solution if it satisfies $f(x) = y$.*

We will identify in the paper several classes of instances of problem (3) for which the relaxation (5) has an optimal E-solution, so in fact it is exact. These classes are distinguished by properties satisfied by the functions f_1, \dots, f_n , and the structure of the constraints in (3). In a first step we group some general facts about problem (3) and its relaxation (5).

2.1 Basic properties of partially separable problems

Although problem (3) is defined for arbitrary functions f_i , we will need to make some assumptions for some of our results. In particular, we will need the following notion of (strong) coercivity.

Definition 2.2. *A function $f : \mathbb{R} \rightarrow \mathbb{R}$ is said to be coercive if $\lim_{|t| \rightarrow +\infty} f(t) = +\infty$, and strongly coercive if $\lim_{|t| \rightarrow +\infty} f(t)/|t| = +\infty$. Moreover, call f semi-coercive if it satisfies at least one of the following two conditions:*

$$\lim_{t \rightarrow -\infty} f(t) = +\infty, \quad \text{or} \quad \lim_{t \rightarrow +\infty} f(t) = +\infty, \quad (6)$$

and strongly semi-coercive if it satisfies at least one of the following two conditions:

$$\lim_{t \rightarrow +\infty} f(t)/t = +\infty, \quad \text{or} \quad \lim_{t \rightarrow -\infty} f(t)/t = -\infty. \quad (7)$$

For instance, any convex non-constant function is semi-coercive. Examples of strongly semi-coercive functions include e^t , e^{-t} , and odd degree univariate polynomials. Examples of strongly coercive functions include $\cosh(t)$ and even degree univariate polynomials with a positive leading coefficient.

When in problem (3) we assume that all the functions f_i are convex, the relaxed problem (5) is convex and thus it admits a dual problem.

Recall that the (Fenchel) conjugate of a function $f : \mathbb{R} \rightarrow \mathbb{R}$ is the function $f^* : \mathbb{R} \rightarrow \mathbb{R} \cup \{\infty\}$ defined by $f^*(y) = \sup_{x \in \mathbb{R}} \{xy - f(x)\}$ for $y \in \mathbb{R}$. When f is strongly coercive, then $f^*(y) \in \mathbb{R}$ for all $y \in \mathbb{R}$.

Theorem 2.1. *The dual problem of problem (5) reads:*

$$\begin{aligned} \sup_{v \in \mathbb{R}^{|L|}} \quad & - \sum_{i=1, \dots, n \mid \beta_{0i} + \sum_{l \in L} \beta_{li} v_l > 0} \left(\beta_{0i} + \sum_{l \in L} \beta_{li} v_l \right) f_i^* \left(- \frac{\alpha_{0i} + \sum_{l \in L} \alpha_{li} v_l}{\beta_{0i} + \sum_{l \in L} \beta_{li} v_l} \right) - \sum_{l \in L} \rho_l v_l \\ \text{s.t.} \quad & v \geq 0, \beta_{0i} + \sum_{l \in L} \beta_{li} v_l \geq 0 \quad (i = 1, \dots, n). \end{aligned} \quad (8)$$

Assume that each function f_i is convex and that problem (5) has a strictly feasible solution. Then the optimal objective value of (5) is equal to the optimal value of the dual problem (8).

Proof. We first compute the Lagrange function $L(x, y, u, v)$ for problem (5), where v_l are the Lagrangian multipliers for the constraints $\alpha_l^T x + \beta_l^T y \leq \rho_l$, and u_i are the multipliers for the constraints $f_i(x_i) - y_i \leq 0$, assumed to be nonnegative. We have:

$$\begin{aligned} L(x, y, u, v) &= \sum_i (\alpha_{0i} x_i + \beta_{0i} y_i) + \sum_i u_i (f_i(x_i) - y_i) + \sum_l (\sum_i (\alpha_{li} x_i + \beta_{li} y_i) - \rho_l) v_l \\ &= \sum_i y_i (\beta_{0i} - u_i + \sum_l \beta_{li} v_l) + \sum_i (x_i (\alpha_{0i} + \sum_l \alpha_{li} v_l) + u_i f_i(x_i)) - \sum_l \rho_l v_l. \end{aligned}$$

Consider the dual objective function $g(u, v) := \inf_{x, y} L(x, y, u, v)$, so that the dual of (5) reads $\sup_{u, v} g(u, v)$. Clearly, $g(u, v) = -\infty$ if, for some i , $\beta_{0i} - u_i + \sum_l \beta_{li} v_l \neq 0$ (letting y_i tend to ∞), or $(u_i = 0$ and $\alpha_{0i} + \sum_l \alpha_{li} v_l \neq 0)$ (letting x_i tend to ∞). Hence, in the dual problem, we can restrict the variables u, v to satisfy

$$\forall i \quad u_i = \beta_{0i} + \sum_l \beta_{li} v_l, \text{ and } \alpha_{0i} + \sum_l \alpha_{li} v_l = 0 \text{ if } u_i = 0.$$

For such u, v , we deduce that

$$\begin{aligned} g(u, v) &= - \sum_{i \mid u_i > 0} \sup_{x_i} x_i (-\alpha_{0i} - \sum_l \alpha_{li} v_l) - u_i f_i(x_i) - \sum_l \rho_l v_l \\ &= - \sum_{i \mid u_i > 0} u_i f_i^* \left(- \frac{\alpha_{0i} + \sum_l \alpha_{li} v_l}{u_i} \right) - \sum_l \rho_l v_l. \end{aligned}$$

This immediately leads to the dual problem (8). The optimal objective values of problems (5) and (8) are equal when (5) is convex and has a strictly feasible solution. \square

For any function $f : \mathbb{R} \rightarrow \mathbb{R}$, the function $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ defined by $h(x, y) = y f^*(x/y)$ for $y > 0$, is a convex function. Hence the objective function in problem (8) is a concave function. Therefore, the dual problem (8) is a maximization optimization problem with linear constraints and a concave objective function.

We now group some results characterizing boundedness of the relaxed problem (5) and providing sufficient conditions for attainment of its minimum; this latter condition plays an important role in several of our results.

Proposition 2.1. *Consider problem (5).*

- (i) *If there does not exist scalars $v_l \geq 0$ ($l \in L$) such that $\beta_0 + \sum_{l \in L} \beta_l v_l \geq 0$, then problem (5) is unbounded (i.e., its minimal value is $-\infty$).*

- (ii) Assume that there exist scalars $v_l \geq 0$ ($l \in L$) such that $\beta_0 + \sum_{l \in L} \beta_l v_l \geq 0$ and that the functions f_i are strongly coercive. Then problem (5) is bounded.
- (iii) Assume that there exist scalars $v_l \geq 0$ ($l \in L$) such that $\beta_0 + \sum_{l \in L} \beta_l v_l > 0$ and that the functions f_i are strongly coercive. Then problem (5) attains its minimum.

Proof. (i) Fix $x \in \mathbb{R}^n$ and consider the following LP in the variables y_1, \dots, y_n :

$$\min \beta_0^T y \quad \text{s.t.} \quad \beta_l^T y \leq \rho_l - \alpha_l^T x \quad (l \in L), \quad y_i \geq f_i(x_i) \quad (i = 1, \dots, n), \quad (9)$$

and its dual LP in the variables v_l ($l \in L$) and u_i ($i = 1, \dots, n$):

$$\max - \sum_{l \in L} v_l (\rho_l - \alpha_l^T x) + \sum_{i=1}^n u_i f_i(x_i) \quad \text{s.t.} \quad \sum_{l \in L} v_l \beta_{li} - u_i = -\beta_{0i} \quad (i = 1, \dots, n), \quad u, v \geq 0. \quad (10)$$

The constraints in (10) can be rewritten as $\sum_{l \in L} v_l \beta_l + \beta_0 \geq 0$ and $v \geq 0$. By assumption, (10) is infeasible, which implies that (9) is unbounded and thus problem (5) too is unbounded.

(ii) By assumption, the dual problem (8) of (5) is feasible. As the functions f_i are strongly coercive, the objective value of (8) at any feasible solution is finite, which thus gives a finite lower bound for the minimum of problem (5). (We use here the weak duality between (5) and (8).)

(iii) By assumption, we have scalars $v_l \geq 0$ ($l \in L$) such that $\beta_0 + \sum_{l \in L} v_l \beta_l > 0$. We show that in problem (5) we can assume without loss of generality that the variables x and y are bounded. Note that each function f_i is bounded from below, since it is strongly coercive. Hence the constraints $y_i \geq f_i(x_i)$ imply that y_i is bounded from below. Therefore, it suffices to show that y_i is bounded from above. For this, pick (x, y) feasible for (5) and write the objective function of (5) as

$$\sum_i \left(\underbrace{(\alpha_{0i} + \sum_l v_l \alpha_{li}) x_i + (\beta_{0i} + \sum_l v_l \beta_{li}) y_i}_{\varphi_i(x_i, y_i)} \right) - \underbrace{\sum_l v_l (\alpha_l^T x + \beta_l^T y)}_{\varphi_0(x, y)} = \sum_i \varphi_i(x_i, y_i) + \varphi_0(x, y).$$

The constraints of (5) imply that the last term satisfies: $\varphi_0(x, y) \geq -\sum_l v_l \rho_l$. As $\beta_{0i} + \sum_l v_l \beta_{li} > 0$, we have that

$$\varphi_i(x_i, y_i) \geq (\alpha_{0i} + \sum_l v_l \alpha_{li}) x_i + (\beta_{0i} + \sum_l v_l \beta_{li}) f_i(x_i)$$

is bounded from below, since the right hand side tends to $+\infty$ when $|x_i|$ tends to $+\infty$ (as f_i is strongly coercive). Since we are minimizing $\alpha_0^T x + \beta_0^T y$, we can assume that it is bounded above and thus we may assume without loss of generality that each variable y_i is bounded and thus x_i too is bounded. This concludes the proof. \square

Remark 2.1. As an application of Proposition 2.1, checking whether the relaxed problem (5) is bounded amounts to checking whether the linear system: $v \geq 0$, $\beta_0 + \sum_{l \in L} v_l \beta_l \geq 0$ is feasible, which can be done with linear programming. Similarly for the linear system: $v \geq 0$, $\beta_0 + \sum_{l \in L} v_l \beta_l > 0$.

2.2 The case of multiple termwise parallel constraints

In this section we consider instances of problem (3), where the functions f_i are assumed not to be dominated by any affine function (see the class \mathcal{F}_1 introduced in Definition 2.4 below) and the m constraints are assumed to be ‘termwise parallel’ - the precise definition is given below. We notice that the analysis in this section also allows that some or all of the constraints in (3) are *equality* constraints.

Definition 2.3. The inequalities: $\alpha_l^T x + \beta_l^T f(x) \leq \rho_l$ ($l \in L$) are termwise parallel if

$$\text{rank}\{(\alpha_{li}, \beta_{li}) : l = 1, \dots, m\} \leq 1 \text{ for all } i = 1, \dots, n, \quad (11)$$

where $\alpha_l = (\alpha_{li})_{i=1}^n$, $\beta_l = (\beta_{li})_{i=1}^n$.

For example, a double-sided inequality: $\rho_1 \leq \sum_{i=1}^n \alpha_i x_i + \beta_i f_i(x_i) \leq \rho_2$ corresponds to two termwise parallel inequalities. Clearly, the constraints: $\alpha_l^T x + \beta_l^T f(x) \leq \rho_l$ ($l \in L$) are termwise parallel if $\alpha_{li} = 0 \forall l \in L \forall i \in \{1, \dots, n\}$.

Example 2.1. As an illustration, the following system:

$$x_1^4 + x_2 + \frac{1}{2}x_2^2 \leq 1, \quad 2x_1^4 - 2x_2 - x_2^2 \leq -1 \quad (12)$$

consists of two termwise parallel constraints. Indeed, choosing the functions $f_1(x_1) = x_1^4$ and $f_2(x_2) = x_2^2$, (12) can be rewritten as $f_1(x_1) + x_2 + \frac{1}{2}f_2(x_2) \leq 1$, $2f_1(x_1) - 2x_2 - f_2(x_2) \leq -1$.

Definition 2.4. Let \mathcal{F}_1 denote the class of continuous functions $f : \mathbb{R} \rightarrow \mathbb{R}$ which are not strictly dominated by an affine function. That is, if $f(t_0) < at_0 + b$ for some $a, b, t_0 \in \mathbb{R}$, then there exists $t_1 \in \mathbb{R}$ such that $f(t_1) = at_1 + b$. Note that $\sup_{\mathbb{R}} f(t) = \infty$ for $f \in \mathcal{F}_1$.

Examples of functions in the class \mathcal{F}_1 include convex non-linear functions and strongly semi-coercive functions.

Our first result shows that when all functions f_i belong to the class \mathcal{F}_1 and the constraints are termwise parallel, if the program (5) has an optimal solution, then it has one which is an E-solution and thus (5) is in fact equivalent to (3).

Theorem 2.2. Consider problem (3), where we assume that each function f_i belongs to the class \mathcal{F}_1 and the m constraints are termwise parallel. If problem (5) attains its minimum, then it has an optimal solution which is an E-solution and thus both problems (3) and (5) are equivalent.

Proof. Let (\bar{x}, \bar{y}) be an optimal solution to problem (5). If (\bar{x}, \bar{y}) is an E-solution, there is nothing to prove. Hence we now assume that the set

$$J := \{i \in \{1, \dots, n\} \mid f_i(\bar{x}_i) < \bar{y}_i\} \quad (13)$$

is not empty. First we show that there exist scalars $\lambda_l \geq 0$ ($l \in L$) satisfying

$$\alpha_{0i} + \sum_{l \in L} \lambda_l \alpha_{li} = 0, \quad \beta_{0i} + \sum_{l \in L} \lambda_l \beta_{li} = 0 \quad \forall i \in J. \quad (14)$$

For this consider the linear program:

$$\begin{aligned} \min \quad & \sum_{i \in J} \alpha_{0i} x_i + \beta_{0i} y_i \quad \text{s.t.} \quad \sum_{i \in J} \alpha_{li} x_i + \beta_{li} y_i \leq \tilde{\rho}_l \quad (l \in L) \\ = \min \quad & a^T x + b^T y \quad \text{s.t.} \quad c_l^T x + d_l^T y \leq \tilde{\rho}_l \quad (l \in L), \end{aligned} \quad (15)$$

and its dual:

$$\max - \sum_{l \in L} \lambda_l \tilde{\rho}_l \quad \text{s.t.} \quad a + \sum_{l \in L} \lambda_l c_l = 0, \quad b + \sum_{l \in L} \lambda_l d_l = 0, \quad \lambda \geq 0 \quad (l \in L), \quad (16)$$

after setting $a = (\alpha_{0i})_{i \in J}$, $b = (\beta_{0i})_{i \in J}$, $c_l = (\alpha_{li})_{i \in J}$, $d_l = (\beta_{li})_{i \in J}$, $\tilde{\rho}_l := \rho_l - (\sum_{i \notin J} \alpha_{li} \bar{x}_i + \beta_{li} \bar{y}_i)$. The linear program (15) is obtained from (5) by ignoring the constraints $f(x) \leq y$ and fixing the variables $x_i = \bar{x}_i$ for $i \notin J$, so that only the variables x_i ($i \in J$) are free. We claim that (16) is

feasible, i.e., the vector $-(a, b)$ lies in the cone generated by the vectors (c_l, d_l) for $l \in L$. If not, Farkas' lemma implies the existence of a vector (u, v) with $a^T u + b^T v > 0$ and $c_l^T u + d_l^T v \geq 0$ for all $l \in L$. Fix a scalar $t > 0$ and define the vectors $x, y \in \mathbb{R}^n$ as follows: For $i \in J$, $x_i = \bar{x}_i - tu_i$, $y_i = \bar{y}_i - tv_i$ and, for $i \notin J$, $x_i = \bar{x}_i$, $y_i = \bar{y}_i$. Then, if we choose t small enough, the inequality $f_i(x_i) < y_i$ holds for all $i \in J$. Hence (x, y) is a feasible solution for (5) whose objective value is smaller than that of (\bar{x}, \bar{y}) , which yields a contradiction. Thus we have shown that there exist nonnegative scalars¹ $\lambda_l \geq 0$ satisfying (14).

Next observe that it suffices to show the existence of \hat{x}_i ($i \in J$) satisfying the conditions:

$$\alpha_{li}\hat{x}_i + \beta_{li}f_i(\hat{x}_i) = \alpha_{li}\bar{x}_i + \beta_{li}\bar{y}_i \quad \text{for all } l \in L, \quad (17)$$

$$\alpha_{0i}\hat{x}_i + \beta_{0i}f_i(\hat{x}_i) = \alpha_{0i}\bar{x}_i + \beta_{0i}\bar{y}_i \quad (18)$$

for all $i \in J$. Indeed, by replacing each \bar{x}_i by \hat{x}_i and \bar{y}_i by $f_i(\hat{x}_i)$ for each $i \in J$, we obtain a new feasible solution for (5) which is an E-solution and whose value is equal to the optimal value of (5). Thus we obtain an optimal solution of (5) which is an E-solution, which will conclude the proof.

To verify (17)-(18) let us fix $i \in J$. Pick an index $l \in L$ for which $(\alpha_{li}, \beta_{li}) \neq (0, 0)$ (if no such index exists then (17) holds for any choice of \hat{x}_i). Consider the function

$$\varphi_{li}(x_i) := \alpha_{li}x_i + \beta_{li}f_i(x_i) - (\alpha_{li}\bar{x}_i + \beta_{li}\bar{y}_i). \quad (19)$$

We claim that there exists \hat{x}_i for which $\varphi_{li}(\hat{x}_i) = 0$. As the constraints are all parallel to the l -th one, this will imply that $\varphi_i^{(l')}(\hat{x}_i) = 0$ for all $l' \in L$ and thus (17) holds. Now, if $\beta_{li} = 0$, just choose $\hat{x}_i = \bar{x}_i$. If $\beta_{li} > 0$, then $\varphi_{li}(\bar{x}_i) = \beta_{li}(f_i(\bar{x}_i) - \bar{y}_i) < 0$. As $f_i \in \mathcal{F}_1$, we deduce the existence of \hat{x}_i for which $\varphi_{li}(\hat{x}_i) = 0$. Analogously if $\beta_{li} < 0$. Hence, for this choice of \hat{x}_i as a root of φ_{li} , condition (17) holds.

We now show that condition (18) holds as well. Indeed, we have

$$\alpha_{0i}\hat{x}_i + \beta_{0i}f_i(\hat{x}_i) = - \sum_l \lambda_l \alpha_{li} \hat{x}_i - \sum_l \lambda_l \beta_{li} f_i(\hat{x}_i) = - \sum_l \lambda_l (\alpha_{li} \bar{x}_i + \beta_{li} \bar{y}_i) = \alpha_{0i} \bar{x}_i + \beta_{0i} \bar{y}_i,$$

where we have used (14) for the first and third equalities and (17) for the second equality. This concludes the proof. \square

Example 2.1 (continued) *We can apply Theorem 2.2 to the problem of minimizing the objective function: $-x_1^4 - 2x_2 - x_2^2$ subject to the constraints (12). Indeed, we are in case (iii) of Proposition 2.1, so that we can conclude that the minimum is attained in the corresponding relaxed problem:*

$$\begin{aligned} \min \quad & -y_1 - 2y_2 \quad \text{s.t.} \quad y_1 + y_2 \leq 1 \\ & 2y_1 - 2y_2 \leq -1 \\ & y_1 \geq x_1^4, \quad y_2 \geq x_2 + \frac{1}{2}x_2^2. \end{aligned}$$

One can easily verify that the optimal solutions of the above relaxed problem are $\bar{x}_1 = \bar{y}_1 = 0$, $\bar{y}_2 = 1$, and $\bar{x}_2 \in [-1 - \sqrt{3}, -1 + \sqrt{3}]$. Now any $\bar{x}_2 \in (-1 - \sqrt{3}, -1 + \sqrt{3})$ is not an E-solution, however two E-solutions are those with $\bar{x}_2 = -1 \pm \sqrt{3}$.

¹Clearly, if we choose a pair (x_i, y_i) ($i \in J$) and λ_l ($l \in L$) of primal/dual optimal solutions to (15), (16), then they satisfy the complementary slackness and thus $\lambda_l = 0$ if the l -th constraint is not active.

Remark 2.2. Consider problem (3), where we assume that the constraints are termwise parallel and that the functions f_i are strongly coercive (a slightly stronger assumption than was made in Theorem 2.2). Then we can assume without loss of generality that the following condition holds for the data of problem (3):

$$\exists i \in \{1, \dots, n\} \text{ such that } \beta_{0i}, \beta_{li} \leq 0 \ (\forall l \in L), \text{ with at least one strict inequality.} \quad (20)$$

Indeed, if (20) does not hold for some index i , then we are in one of the following ‘pathological’ cases:

- (i) $\beta_{0i} < 0$. Then (5) is unbounded (by Proposition 2.1 (i)), and problem (3) as well if $|L| = 1$.
- (ii) $\beta_{0i} = 0$ and (say) $\beta_{1i} < 0$, so that $\beta_{li} = \lambda_l \beta_{1i}$, $\alpha_{li} = \lambda_l \alpha_{1i}$ (for some scalar λ_l) for all $l \in L$, and thus $\beta_{li} = 0$ implies $\alpha_{li} = 0$ (since the constraints are parallel). If $\alpha_{0i} \neq 0$, then (3) is unbounded (since we can let $\alpha_{0i} x_i$ tend to $-\infty$). If $\alpha_{0i} = 0$, then any constraint with $\beta_{li} < 0$ is redundant and can thus be eliminated (and variable x_i can be eliminated as well, so we can reduce to a problem with less variables and constraints).

Note that, in case (i), the relaxed problem (5) is unbounded, although the original (3) might be bounded when $|L| \geq 2$. This situation is illustrated by the following small example: $\min_{0 \leq x \leq 1} x - x^2$, with two constraints and with optimal value 0, while the relaxed problem: $\min_{0 \leq x \leq 1, y \geq x^2} x - y$ is unbounded.

Extension to the case of block-separable variables

Theorem 2.2 extends directly to the more general setting where, in problem (3), we replace single variables by disjoint blocks of variables. More precisely, we assume that the variables x_1, \dots, x_n are partitioned into N disjoint blocks of variables $x_{[1]}, \dots, x_{[N]}$, where $[1], \dots, [N]$ denote subsets forming a partition of $\{1, \dots, n\}$ and $x_{[j]} = (x_i \mid i \in [j])$ denotes the group of variables indexed by $[j]$. We now assume that we have functions f_1, \dots, f_N , where f_j is a function from $\mathbb{R}^{|[j]|}$ to \mathbb{R} , and parameters $\alpha_l \in \mathbb{R}^n$, $\beta_{lj} \in \mathbb{R}$ for $j \in \{1, \dots, N\}$ and $l \in \{0\} \cup L$. We consider the following extension of problem (3):

$$\min \alpha_0^T x + \sum_{j=1}^N \beta_{0j} f_j(x_{[j]}) \quad \text{s.t.} \quad \alpha_l^T x + \sum_{j=1}^N \beta_{lj} f_j(x_{[j]}) \leq \rho_l \quad (l \in L), \quad (21)$$

which is still a special case of problem (NCP), and its relaxation:

$$\begin{aligned} \min \alpha_0^T x + \sum_{j=1}^N \beta_{0j} y_j \quad \text{s.t.} \quad & \alpha_l^T x + \sum_{j=1}^N \beta_{lj} y_j \leq \rho_l \quad (l \in L) \\ & f_j(x_{[j]}) \leq y_j \quad (j \in \{1, \dots, N\}). \end{aligned} \quad (22)$$

The definitions of termwise parallel constraints and of the class of functions \mathcal{F}_1 (from Definitions 2.3 and 2.4) extend to the multivariate case as follows.

Definition 2.5. The inequalities $\alpha_l^T x + \sum_{j=1}^N \beta_{lj} f_j(x_{[j]}) \leq \rho_l$ ($l \in L$) are termwise parallel if

$$\text{rank}\{(\alpha_{li} \ (i \in [j]), \beta_{lj}) : l \in L\} \leq 1 \text{ for all } j \in \{1, \dots, N\}. \quad (23)$$

Definition 2.6. A continuous function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ belongs to \mathcal{F}_1 if there is no affine linear d -variate polynomial ℓ for which $f(t) < \ell(t)$ for all $t \in \mathbb{R}^d$.

For instance, the function $x \mapsto \|x\|_p$ (for $p > 0$) belongs to \mathcal{F}_1 .

Example 2.2. *As an illustration, consider the following small example:*

$$\begin{aligned} \min \quad & x_1 - x_2 + x_3 + 2x_4 + (x_1^2 - x_1x_2) + 2(x_3^2 + 3x_4^2 - x_3x_4) \\ \text{s.t.} \quad & 2x_1 + 3(x_1^2 - x_1x_2) - x_4 - (x_3^2 + 3x_4^2 - x_3x_4) \leq 2 \\ & -4x_1 - 6(x_1^2 - x_1x_2) + 2x_4 + 2(x_3^2 + 3x_4^2 - x_3x_4) \leq -1. \end{aligned} \quad (24)$$

Here, we have two blocks of variables $x_{[1]} = (x_1, x_2)$ and $x_{[2]} = (x_3, x_4)$, we use the functions $f_1(x_1, x_2) = x_1^2 - x_1x_2$ and $f_2(x_3, x_4) = x_3^2 + 3x_4^2 - x_3x_4$, and the two constraints are termwise parallel.

We have the following analogue of Theorem 2.2.

Theorem 2.3. *Consider problem (21), where we assume that the functions f_1, \dots, f_N belong to the class \mathcal{F}_1 and that the constraints are termwise parallel. If problem (22) attains its minimum, then it has an optimal E-solution and thus both problems (21) and (22) are equivalent.*

Proof. Let (\bar{x}, \bar{y}) be an optimal solution to (22) and set $J = \{j \in \{1, \dots, N\} \mid f_j(\bar{x}_{[j]}) < \bar{y}_j\}$. As in the proof of Theorem 2.2, there exist multipliers $\lambda_l \geq 0$ ($l \in L$) such that

$$\alpha_{0i} + \sum_{l \in L} \lambda_l \alpha_{li} = 0 \quad \forall i \in \bigcup_{j \in J} [j], \quad \beta_{0j} + \sum_{l \in L} \lambda_l \beta_{lj} = 0 \quad \forall j \in J. \quad (25)$$

Define the function

$$\varphi_{lj}(x_{[j]}) = \sum_{i \in [j]} \alpha_{li}(x_i - \bar{x}_i) + \beta_{lj}(f_j(x_{[j]}) - \bar{y}_j) \quad (26)$$

for $l \in L$, $j \in J$. Pick $l_1 \in L$ for which the vector $(\alpha_{l_1 i} \ (i \in [j]), \beta_{l_1 j})$ is not identically zero. As the constraints are termwise parallel, any φ_{lj} is a scalar multiple of $\varphi_{l_1 j}$. Now, as in the proof of Theorem 2.2, it suffices to show the existence of $\hat{x}_{[j]}$ ($j \in J$) for which $\varphi_{l_1 j}(\hat{x}_{[j]}) = 0$ for all $j \in J$. This follows using the assumption that $f_j \in \mathcal{F}_1$; we omit the details which are analogous to those in the proof of Theorem 2.2. \square

As was done for problem (5), one can write explicitly the dual problem of problem (22), using the Fenchel conjugate functions f_j^* . For a multivariate function $f : \mathbb{R}^d \rightarrow \mathbb{R}$, its conjugate is the function $f^* : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}$ defined by $f^*(y) = \sup_{x \in \mathbb{R}^d} \{x^T y - f(x)\}$ for $y \in \mathbb{R}^d$. Again, $f^*(y) \in \mathbb{R}$ for all $y \in \mathbb{R}^d$ when f is strongly coercive. The definitions for the univariate case from Definition 2.2 extend to the multivariate setting as follows.

Definition 2.7. *A function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is strongly semi-coercive if it satisfies the following analogue of condition (7):*

$$\frac{f(x)}{\|x\|_2} \rightarrow +\infty \text{ on at least one half-line of } \mathbb{R}^d \quad (27)$$

and f is strongly coercive if $\lim_{\|x\|_2 \rightarrow +\infty} f(x)/\|x\|_2 = +\infty$.

The following analogue of Theorem 2.1 holds; we omit the proof since it is analogous to that of Theorem 2.1.

Theorem 2.4. *Consider problem (22), where we assume that each function f_j is convex. If problem (22) has a strictly feasible solution, then its optimal objective value is equal to the optimal value of the following dual problem:*

$$\begin{aligned} \sup_{v \in \mathbb{R}^{|L|}} \quad & - \sum_{j=1, \dots, N \mid \beta_{0j} + \sum_j \beta_{lj} v_l > 0} (\beta_{0j} + \sum_{l=1}^N \beta_{lj} v_l) f_j^* \left(- \frac{(\alpha_{0i} + \sum_{l \in L} \alpha_{li} v_l)_{i \in [j]}}{\beta_{0j} + \sum_{l=1}^N \beta_{lj} v_l} \right) - \sum_{l \in L} \rho_l v_l \\ \text{s.t.} \quad & v_l \geq 0 \quad \forall l \in L, \quad \beta_{0j} + \sum_{l=1}^N \beta_{lj} v_l \geq 0 \quad \forall j = 1, \dots, N. \end{aligned} \quad (28)$$

2.3 The case of two constraints

In the preceding section we have considered instances of problem (3), where all constraints are assumed to be termwise parallel and the functions f_i belong to the class \mathcal{F}_1 . If we consider stronger properties for the functions f_i (see the class \mathcal{F}_2 in Definition 2.8 below), then we can allow two arbitrary constraints and still show a hidden convexity type result. We will also consider some extensions where we allow more constraints (but satisfying some parallelism conditions).

Definition 2.8. *Let \mathcal{F}_2 denote the class of continuous univariate functions f satisfying the following condition: If $f(t_0) < at_0 + b$ for some $a, b, t_0 \in \mathbb{R}$, then there exist scalars $t_1, t_2 \in \mathbb{R}$ such that $t_1 < t_0 < t_2$, $f(t_1) = at_1 + b$, and $f(t_2) = at_2 + b$.*

Examples of functions in the class \mathcal{F}_2 include coercive convex functions (e.g., e^{t^2} , t^4 , $\cosh(t)$) and strongly coercive functions (e.g., even degree polynomials with a positive leading coefficient).

Theorem 2.5. *Consider problem (3) with $m = 2$ constraints and where the functions f_1, \dots, f_n are assumed to belong to the class \mathcal{F}_2 . Assume that problem (5) has an optimal solution (\bar{x}, \bar{y}) . Then there exist scalars $\lambda_1, \lambda_2 \geq 0$ satisfying*

$$\alpha_{0i} + \lambda_1 \alpha_{1i} + \lambda_2 \alpha_{2i} = 0, \quad \beta_{0i} + \lambda_1 \beta_{1i} + \lambda_2 \beta_{2i} = 0 \quad \forall i \in J := \{i \mid f_i(\bar{x}_i) < \bar{y}_i\}. \quad (29)$$

Assume, moreover, that $\lambda_1 \lambda_2 = 0$ and that the following condition holds:

$$\text{There does not exist } i \in J \text{ such that } \beta_{0i}, \beta_{1i}, \beta_{2i} \leq 0, \text{ with at least one strict inequality.} \quad (30)$$

Then problem (5) has an optimal E-solution and thus both problems (3) and (5) are equivalent.

Proof. The proof is along the same line as for Theorem 2.2, with some additional technical details. Let (\bar{x}, \bar{y}) be an optimal solution of (5) and set $J := \{i \mid f_i(\bar{x}_i) < \bar{y}_i\}$. We assume that $J \neq \emptyset$ (else we are done). The existence of $\lambda_1, \lambda_2 \geq 0$ satisfying (29) was shown in the first part of the proof of Theorem 2.2 (recall (14) there). By assumption, $\lambda_1 \lambda_2 = 0$. Say, $\lambda_2 = 0$, so that (29) implies:

$$\alpha_{0i} + \lambda_1 \alpha_{1i} = 0, \quad \beta_{0i} + \lambda_1 \beta_{1i} = 0 \quad \forall i \in J. \quad (31)$$

As in the proof of Theorem 2.2, we consider the functions $\varphi_{li}(x_i)$ ($l = 1, 2$) from (19). It suffices to show that we can find \hat{x}_i satisfying (18) and

$$\varphi_{1i}(\hat{x}_i) = 0, \quad \varphi_{2i}(\hat{x}_i) \leq 0 \quad \forall i \in J. \quad (32)$$

Indeed, replacing (\bar{x}_i, \bar{y}_i) by $(\hat{x}_i, f_i(\hat{x}_i))$ for $i \in J$, we obtain an optimal solution of (5) which is an E-solution, thus showing equivalence of (3) and (5).

Suppose first that $\beta_{1i} > 0$. Then $\varphi_{1i}(\bar{x}_i) = \beta_{1i}(f_i(\bar{x}_i) - \bar{y}_i) < 0$. As $f_i \in \mathcal{F}_2$, the function φ_{1i} has two roots r_i and s_i such that $r_i < \bar{x}_i < s_i$. For any root t of φ_{1i} , one has that $\varphi_{2i}(t) = (t - \bar{x}_i) \left(\alpha_{2i} - \frac{\beta_{2i} \alpha_{1i}}{\beta_{1i}} \right)$. Thus we can choose $t \in \{r_i, s_i\}$ such that $\varphi_{2i}(t) \leq 0$. Denoting by \hat{x}_i this choice of the root of φ_{1i} , we thus have found \hat{x}_i satisfying (32). The same conclusion holds assuming $\beta_{1i} < 0$.

Suppose now that $\beta_{1i} = 0$. If $\alpha_{1i} = 0$, then φ_{1i} is identically zero and we can choose for \hat{x}_i a root of φ_{2i} . If $\alpha_{1i} \neq 0$, then we must choose $\hat{x}_i = \bar{x}_i$ as this is the only root of φ_{1i} . Then $\varphi_{2i}(\bar{x}_i) = \beta_{2i}(f_i(\bar{x}_i) - \bar{y}_i)$, so that (32) holds if $\beta_{2i} \geq 0$. Hence we are left with the case when $\beta_{1i} = 0, \beta_{2i} < 0, \beta_{0i} = 0$ (by (31)), which is excluded by assumption (30).

Finally we verify that condition (18) holds. Using (31) combined with the fact that $\varphi_{1i}(\hat{x}_i) = 0$ from (32), we obtain:

$$\alpha_{0i}\hat{x}_i + \beta_{0i}f_i(\hat{x}_i) = -\lambda_1(\alpha_{1i}\hat{x}_i + \beta_{1i}f_i(\hat{x}_i)) = \alpha_{0i}\bar{x}_i + \beta_{0i}\bar{y}_i,$$

thus showing (18) and concluding the proof. \square

Remark 2.3. We make some comments regarding condition (30), which is related to the condition (20) introduced earlier in Remark 2.2 following Theorem 2.2. Again, if we assume that the functions f_i are strongly coercive, then we can assume without loss of generality that the data of problem (3) satisfy the condition (30). Indeed suppose that condition (30) does not hold, i.e., there exists an index i such that $\beta_{0i}, \beta_{1i}, \beta_{2i} \leq 0$, with at least one strict inequality. Then one can verify that

- (i) either problem (3) is unbounded,
- (ii) or at least one of the two constraints is redundant,
- (iii) or Theorem 2.5 does not hold, i.e., problem (3) is not equivalent to (5).

Case (i) is covered by Proposition 2.1 (i). Case (ii) occurs when $\alpha_{0i} = \beta_{0i} = 0$. Case (iii) occurs in the case when $\beta_{0i} = 0$ and one of β_{1i}, β_{2i} is zero, say $\beta_{1i} = 0$, and $\beta_{2i} < 0$, $\alpha_{0i}\alpha_{1i} < 0$. Example 2.3 below illustrates case (iii).

Example 2.3. We give here an example showing that if $\lambda_1\lambda_2 = 0$ does hold, but condition (30) does not hold, then the conclusion of Theorem 2.5 does not hold. Consider the following instance of problem (3) with two constraints:

$$\pi^* := \min -x_1 + x_2 + x_2^2 \quad \text{s.t.} \quad \begin{aligned} x_1 + x_2 + x_2^2 &\leq 1 \\ x_1 - x_1^2 + x_2 - x_2^2 &\leq -\frac{3}{2}, \end{aligned} \quad (33)$$

and its relaxation

$$\pi := \min -x_1 + x_2 + y_2 \quad \text{s.t.} \quad \begin{aligned} x_1 + x_2 + y_2 &\leq 1 \\ x_1 - y_1 + x_2 - y_2 &\leq -\frac{3}{2} \\ y_1 &\geq x_1^2, \quad y_2 \geq x_2^2. \end{aligned} \quad (34)$$

In a first step, we show that the optimal value of (34) is $\pi = -\frac{3}{2}$ and is attained at $x_1 = \frac{5}{4}$, $y_1 \geq 2$ ($> x_1^2$), $x_2 = -\frac{1}{2}$, $y_2 = \frac{1}{4}$ ($= x_2^2$); we also show that the second multiplier satisfies $\lambda_2 = 0$.

To see this, let (x, y) be feasible for (34). Observe that $x_2 + y_2 \geq x_2 + x_2^2 \geq -\frac{1}{4}$, with equality if $x_2 = -\frac{1}{2}$, $y_2 = x_2^2 = \frac{1}{4}$. This fact combined with the first inequality in (34) implies:

$$-x_1 + x_2 + y_2 = -(x_1 + x_2 + y_2) + 2(x_2 + y_2) \geq -1 - \frac{1}{2} = -\frac{3}{2}.$$

Therefore, $\pi \geq -\frac{3}{2}$. Moreover, equality $\pi = -\frac{3}{2}$ holds precisely when $x_2 = -\frac{1}{2}$, $y_2 = x_2^2 = \frac{1}{4}$, $x_1 = \frac{5}{4}$ and, using the second constraint in (34), $y_1 \geq 2$.

Hence $J = \{1\}$. Note that, if we choose $y_1 = 2$, then the two constraints are active. Moreover, the relations for the multipliers read: $-1 + \lambda_1 + \lambda_2 = 0$ and $0 + 0\lambda_1 - \lambda_2 = 0$, implying $\lambda_2 = 0$ and $\lambda_1 = 1$.

Next we verify that the optimal value of (33) satisfies $\pi^* > \pi = -\frac{3}{2}$. For this assume that there exist x_1, x_2 satisfying

- (a) $-x_1 + x_2 + x_2^2 = -\frac{3}{2}$,
- (b) $x_1 + x_2 + x_2^2 \leq 1$
- (c) $x_1 - x_1^2 + x_2 - x_2^2 \leq -\frac{3}{2}$.

Relation (a) implies $-x_1 + (x_2 + \frac{1}{2})^2 + \frac{5}{4} = 0$ and thus $x_1 \geq \frac{5}{4}$. Summing (a) and (b) we get: $x_2 + x_2^2 \leq -\frac{1}{4}$, i.e., $(x_2 + \frac{1}{2})^2 \leq 0$. implying $x_2 = -\frac{1}{2}$. Using (a), this implies $x_1 = \frac{5}{4}$. Therefore, $x_1 - x_1^2 + x_2 - x_2^2 = \frac{5}{4} - \frac{25}{16} - \frac{1}{2} - \frac{1}{4} = \frac{1}{2} - \frac{25}{16} > -\frac{3}{2}$, thus contradicting (c).

In this example, the parameters satisfy: $\beta_{0i} = \beta_{1i} = 0$, $\beta_{2i} < 0$, $\alpha_{0i}\alpha_{1i} < 0$, so that condition (30) is violated. The relaxed program (34) has an optimal solution where the two constraints are active and the second multiplier is $\lambda_2 = 0$. Yet the relaxed problem (34) is not equivalent to the original problem (33), which verifies the statement in Remark 2.3 (iii). Hence the result of Theorem 2.5 does not hold in this case.

Extension to the case of block-separable variables and multiple constraints

The result of Theorem 2.5 extends to the case where we have block-separable variables and multiple constraints of a special form as specified below.

Consider problem (21) involving the block-separable variables $x_{[j]}$ ($j = 1, \dots, N$) and its relaxation (22). We assume that we have two groups of constraints. The constraints in the first group, indexed by L_1 , are termwise parallel, i.e., satisfy (23). The constraints in the second group, indexed by L_2 , are also termwise parallel and, moreover, satisfy the following stronger sign condition:

$$\forall j \in \{1, \dots, N\} \exists l_2 \in L_2 \forall l \in L_2 \exists \mu_l \geq 0 \text{ such that } \alpha_{li} = \mu_l \alpha_{l_2 i} \forall i \in [j] \text{ and } \beta_{lj} = \mu_l \beta_{l_2 j}. \quad (35)$$

We will make the following assumption on the parameters β_0, β_l (which can be seen as an analogue of condition (30) in the separable case):

$$\begin{aligned} \exists j \in \{1, \dots, N\} \text{ such that } \beta_{0j} = 0, \beta_{lj} = 0 \forall l \in L_1, \\ \text{and } \beta_{lj} \leq 0 \forall l \in L_2, \text{ with at least one strict inequality.} \end{aligned} \quad (36)$$

Under these conditions the following extension of Theorem 2.5 holds.

Theorem 2.6. *Consider problem (21). We assume that condition (36) holds as well as the following two conditions:*

- (i) *When each block $[j]$ has cardinality 1 (the usual separable case), we assume that $f_j \in \mathcal{F}_2$ for all j ; otherwise we assume that the functions f_1, \dots, f_N are strongly coercive.*
- (ii) *The constraints are indexed by $L = L_1 \cup L_2$, where the constraints indexed by L_1 are termwise parallel and those indexed by L_2 satisfy (35).*

Assume that problem (22) attains its minimum at (\bar{x}, \bar{y}) . Then there exist scalars $\lambda_l \geq 0$ ($l \in L_1 \cup L_2$) satisfying

$$\alpha_{0i} + \sum_{l \in L_1 \cup L_2} \lambda_l \alpha_{li} = 0 \forall i \in \bigcup_{j \in J} [j], \beta_{0j} + \sum_{l \in L_1 \cup L_2} \lambda_l \beta_{lj} = 0 \forall j \in J, \quad (37)$$

where $J := \{j \in \{1, \dots, N\} \mid f_j(\bar{x}_{[j]}) < \bar{y}_j\}$. If, moreover, $\lambda_l = 0$ for all $l \in L_2$, then problem (22) is equivalent to problem (21).

Proof. The proof combines ideas from the proofs of Theorems 2.3 and 2.5. We only point out the main steps. Let $j \in J$. Pick $l_1 \in L_1$ for which the vector $(\alpha_{l_1 i} \ (i \in [j]), \beta_{l_1 j})$ is not identically zero; pick analogously such $l_2 \in L_2$. Recall the definition of the function $\varphi_{l_j}(x_{[j]})$ from (26). By the assumption on the constraints, $\varphi_{l_j}(x_{[j]})$ is a scalar multiple of $\varphi_{l_1 j}(x_{[j]})$ for any $l \in L_1$, and

$\varphi_{l_j}(x_{[j]})$ is a nonnegative scalar multiple of $\varphi_{l_{2j}}(x_{[j]})$ for any $l \in L_2$. Hence it suffices to show the existence of $\hat{x}_{[j]}$ such that

$$\varphi_{l_{1j}}(\hat{x}_{[j]}) = 0 \quad \text{and} \quad \varphi_{l_{2j}}(x_{[j]}) \leq 0. \quad (38)$$

Suppose first that $\beta_{l_{1j}} \neq 0$. Then, for any vector $x_{[j]}$,

$$\varphi_{l_{1j}}(x_{[j]}) = 0 \implies \varphi_{l_{2j}}(x_{[j]}) = \sum_{i \in [j]} \underbrace{\left(\alpha_{l_{2i}} - \beta_{l_{2i}} \frac{\alpha_{l_{1i}}}{\beta_{l_{1i}}} \right)}_{=: \sigma_i} (x_i - \bar{x}_i). \quad (39)$$

Let ϵ_i denote the sign of the parameter σ_i appearing in the above relation (39). Say $\beta_{l_{1j}} > 0$ (the case $\beta_{l_{1j}} < 0$ is similar). Then, $\varphi_{l_{1j}}(\bar{x}_{[j]}) = \beta_{l_{1j}}(f_j(\bar{x}_{[j]}) - \bar{y}_j) < 0$. For $t \geq 0$, define the vector $x_{[j]}$, where

$$x_i = \bar{x}_i - \epsilon_i t \quad \text{for} \quad i \in [j].$$

Letting $t \rightarrow +\infty$, the norm of $x_{[j]}$ tends to $+\infty$ and thus $\varphi_{l_{1j}}(x_{[j]})$ tends to $+\infty$ (since f_j is strongly coercive). Therefore, there must exist $t \geq 0$ for which $\varphi_{l_{1j}}(x_{[j]}) = 0$ and thus, in view of (39), $\varphi_{l_{2j}}(x_{[j]}) \leq 0$. This shows (38).

Suppose now $\beta_{l_{1j}} = 0$. This implies $\beta_{l_j} = 0$ for all $l \in L_1$ (since the constraints are termwise parallel) and, in turn, this implies $\beta_{0j} = 0$ (using (37)). Finally, $\beta_{l_{2j}} \geq 0$; indeed, using (35), $\beta_{l_{2j}} < 0$ would imply $\beta_{l_j} \leq 0$ for all $l \in L_2$, which contradicts (36). Now, choose $\hat{x}_{[j]} = \bar{x}_{[j]}$, so that $\varphi_{l_{1j}}(\bar{x}_{[j]}) = 0$ and $\varphi_{l_{2j}}(\bar{x}_{[j]}) = \beta_{l_{2j}}(f_j(\bar{x}_{[j]}) - \bar{y}_j) \leq 0$, thus showing (38).

When all blocks $[j]$ have cardinality 1, we can show the existence of $\hat{x}_{[j]}$ satisfying (38) using only the assumption $f_j \in \mathcal{F}_2$, in the same way as in the proof of Theorem 2.5. This concludes the proof. \square

We conclude this section with Table 1, which summarizes the problem instances treated in this section and the previous one, where we could prove a hidden convexity result.

2.4 Application to quadratic optimization

We discuss here some problems, arising in particular from quadratic optimization, which can be cast as instances of the separable problems treated in the previous sections.

Consider a general quadratic optimization problem:

$$\min a^T x + x^T B x \quad \text{s.t.} \quad c_l^T x + x^T D_l x \leq \rho_l \quad (l \in L) \quad (40)$$

where B, D_l are real symmetric $n \times n$ matrices, $a, c_l \in \mathbb{R}^n$, $\rho_l \in \mathbb{R}$.

Definition 2.9. *Two matrices A, B are said to be simultaneously diagonalizable (SD) if there exists a nonsingular matrix U such that both $U^T A U$ and $U^T B U$ are diagonal matrices.*

As was observed in [3], when the matrices B, D_l ($l \in L$) can be simultaneously diagonalized, problem (40) can be cast as an instance of problem (3), where we choose the functions $f_i(x_i) = x_i^2$. Indeed, if $U^T B U, U^T D_l U$ are all diagonal, then we can make a linear change of variables (replacing x by $U^{-1}x$), and reformulate (40) as

$$\min \alpha_0^T x + \sum_{i=1}^n \beta_{0i} x_i^2 \quad \text{s.t.} \quad \alpha_l^T x + \sum_{i=1}^n \beta_{li} x_i^2 \leq \rho_l \quad (l \in L), \quad (41)$$

Constraints	Conditions	Hidden convexity results
$ L \geq 1$	<ul style="list-style-type: none"> • $f_i \in \mathcal{F}_1$ (Definition 2.4) • Problem (5) has an optimal solution • Termwise parallel constraints if $L > 1$ (Definition 2.3) 	Theorem 2.2 (separable case)
$ L \geq 1$	<ul style="list-style-type: none"> • $f_i \in \mathcal{F}_1$ (Definition 2.6) • Problem (22) has an optimal solution • Termwise parallel constraints if $L > 1$ (Definition 2.5) 	Theorem 2.3 (block-separable case)
$ L = 2$	<ul style="list-style-type: none"> • $f_i \in \mathcal{F}_2$ (Definition 2.8) • Problem (5) has an optimal solution • One of the dual multipliers is zero • The sign condition (30) holds 	Theorem 2.5 (separable case)
$L = L_1 \cup L_2$	<ul style="list-style-type: none"> • $f_i \in \mathcal{F}_2$ (separable case) • f_i is strongly coercive (block-separable case) • Problem (22) has an optimal solution • The constraints in L_1 are termwise parallel • The constraints in L_2 satisfy (35) • The dual multipliers of the constraints in L_2 are equal to 0 • The sign condition (36) holds 	Theorem 2.6 (separable and block-separable cases)

Table 1: Summary of the results in Sections 2.2 and 2.3

where $\alpha_0 = Ua$, $\alpha_l = Uc_l$, $U^T B U = \text{diag}(\beta_1, \dots, \beta_n)$, $U^T D_l U = \text{diag}(\beta_{l1}, \dots, \beta_{ln})$. The corresponding relaxation (5) of problem (41) reads:

$$\min \alpha_0^T x + \beta_0^T y \text{ s.t. } \alpha_l^T x + \beta_l^T y \leq \rho_l \ (l \in L), \ y_i \geq x_i^2 \ (i = 1, \dots, n). \quad (42)$$

First let us observe that this relaxation coincides in fact with the basic semidefinite programming relaxation of the quadratic problem (41). Indeed, the basic SDP relaxation has additional variables z_{ij} (for the pairwise products $x_i x_j$) and, instead of the conic quadratic constraints: $y_i \geq x_i^2$, it requires that the matrix

$$Y := \begin{pmatrix} 1 & x_1 & x_2 & \dots & x_n \\ x_1 & y_1 & z_{12} & \dots & z_{1n} \\ x_2 & z_{12} & y_2 & \dots & z_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_n & z_{1n} & z_{2n} & \dots & y_n \end{pmatrix}$$

be positive semidefinite. Note that $Y \succeq 0$ implies $y_i \geq x_i^2$ for all i which, in turn, implies the existence of scalars z_{ij} ($1 \leq i < j \leq n$) achieving $Y \succeq 0$. As the monomials $x_i x_j$ do not appear in (41), we can conclude that (42) coincides with the basic SDP relaxation of (41).

Next we note that, as an application of Proposition 2.1, the relaxed problem (42) is bounded if and only if there exists $v \geq 0$ such that $B + \sum_l v_l D_l \succeq 0$ and, moreover, if $B + \sum_l v_l D_l \succ 0$ for some $v \geq 0$, then (42) attains its minimum.

Simultaneous diagonalization

In order to bring problem (40) into the separable form (41), we need to be able to simultaneously diagonalize the matrices B, D_l ($l \in L$).

As is well known, the matrices B, D_l ($l \in L$) are SD if they pairwise commute. A first obvious case when this is possible is when $D_l = 0$ for all l , corresponding to the case when all constraints in problem (40) are linear. Then, after diagonalizing B , one obtains the reformulation (41), where all constraints are linear (i.e., $\beta_l = 0$ for all $l \in L$). Hence the constraints are termwise parallel and we are thus in the setting of Theorem 2.2. However, as observed above, the relaxed problem (42) is bounded if and only if $B \succeq 0$, which is precisely the case when the original problem (40) (or (41)) is already convex, so that using the relaxation (42) does not help. As an illustration, recall that the maximum cut problem can be formulated as the instance:

$$\min -x^T Lx \quad \text{s.t. } x \in [-1, 1]^n \quad (43)$$

of problem (40), thus with linear constraints. Here, L is the Laplacian matrix of the graph which is known to be positive semidefinite. The fact that we cannot apply Theorem 2.2 to reformulate (43) in the form (42) is consistent with the fact that the maximum cut problem is NP-hard.

Another useful sufficient condition for simultaneous diagonalization of *two* matrices is provided by the following classical result (proved e.g. in [11]).

Theorem 2.7. *Given two real symmetric $n \times n$ matrices A and B , let $Q_A = \{x \mid x^T A x = 0\}$ and $Q_B = \{x \mid x^T B x = 0\}$. If the condition:*

$$Q_A \cap Q_B = \{0\} \quad (44)$$

holds, then A and B can be simultaneously diagonalized. Moreover, when $n \geq 3$, condition (44) holds if and only if there exists a definite matrix in the pencil $\mathbb{R}(A, B) = \{\alpha A + \beta B \mid \alpha, \beta \in \mathbb{R}\}$ (and then all matrices in the pencil are SD).

Hence, two matrices are SD as soon as one of them is definite. Moreover, if, in the quadratic optimization problem (40), the matrices B, D_l ($l \in l$) are all linear combinations of two matrices, one of them definite, then the problem can be reformulated as an instance of (41). This is the case, e.g., when $Q_B \cap Q_{D_1} = \{0\}$ and we are in one of the following cases:

- (i) (40) has only one constraint,
- (ii) (40) has two constraints, one of them linear (i.e., $D_2 = 0$),
- (iii) (40) has a double-sided constraint (i.e., $D_2 = -D_1$).

The next result summarizes what we can claim when the quadratic problem (40) has one or two constraints as in cases (i), (ii), (iii) above.

Proposition 2.2. *Consider problem (40) assumed to have one constraint, or two constraints in which case we assume that $D_2 = \lambda D_1$ for some $\lambda \in \mathbb{R}$. Assume that there exists $v_1 \geq 0$ such that $B + v_1 D_1 \succ 0$. Then the matrices B and D_1 are SD, hence one can reformulate (40) as the quadratic (in general non-convex) separable problem (41) which, in turn, is equivalent to the convex problem (42).*

The following example, taken from [3], illustrates the above case (ii) where one of the two constraints is linear. Namely, if an optimization problem has a quadratic objective function and a conic quadratic constraint of the form:

$$\|Ax + b\|_2 \leq c^T x + d,$$

then this constraint can be reformulated as a quadratic constraint combined with a linear constraint:

$$\|Ax + b\|^2 \leq (c^T x + d)^2, \quad c^T x + d \geq 0,$$

which thus falls within the above case (ii).

Possible extension

By Theorem 2.7, two symmetric matrices A and B can be simultaneously diagonalized when the set $Q_A \cap Q_B$ of common real zeros of their quadratic forms is reduced to the zero vector. To conclude let us mention the following possible extension to the case when $Q_A \cap Q_B$ has dimension at least 1.

Lemma 2.1. *If A, B are symmetric $n \times n$ matrices with $\dim(Q_A \cap Q_B) = k \geq 1$, then there exists a nonsingular matrix U such that both $U^T A U$ and $U^T B U$ have an arrow shape like:*

$$\begin{pmatrix} X & * & \dots & * \\ * & * & & \\ \vdots & & \ddots & \\ * & & & * \end{pmatrix}, \quad (45)$$

where X is any $k \times k$ matrix with zero diagonal entries and the stars $*$ indicate blocks of positions where the entries may be nonzero (thus all entries are equal to 0 at positions (i, i) for $1 \leq i \leq k$, and at positions (i, j) for $k + 1 \leq i \neq j \leq n$).

Proof. Say $Q_A \cap Q_B$ is spanned by k linearly independent vectors u_1, \dots, u_k . Up to a change of base we can assume that u_1, \dots, u_k are the first k unit vectors e_1, \dots, e_k , so that $A_{ii} = B_{ii} = 0$ for $i = 1, \dots, k$. Let A', B' be the principal submatrices of A, B obtained by deleting their first k rows and columns. Then, $Q_{A'} \cap Q_{B'} = \{0\}$. Indeed, if u is a nonzero vector in $Q_{A'} \cap Q_{B'}$, then together with $\{e_1, \dots, e_k\}$, this would give $k + 1$ linearly independent vectors in $Q_A \cap Q_B$. Hence, we can apply Theorem 2.7 and obtain a nonsingular matrix W such that both $W^T A' W$ and $W^T B' W$ are diagonal. Define the matrix $U = \begin{pmatrix} I_k & 0 \\ 0 & W \end{pmatrix}$. Then U is nonsingular and both $U^T A U$ and $U^T B U$ have the desired arrow shape (45). \square

Consider, for instance, problem (40) having just one constraint and where $Q_B \cap Q_{D_1}$ has dimension $k = 1$. Then, we can apply the above lemma and derive a reformulation of the form:

$$\min \alpha_0^T x + \sum_{i=2}^n \beta_{0i} x_i^2 + x_1 \left(\sum_{i=2}^n \beta'_{0i} x_i \right) \quad \text{s.t.} \quad \alpha_1^T x + \sum_{i=2}^n \beta_{1i} x_i^2 + x_1 \left(\sum_{i=2}^n \beta'_{1i} x_i \right) \leq \rho.$$

Here a strategy is to do binary search on x_1 . For each fixed x_1 , we obtain the following problem in separable form:

$$\begin{aligned} g_0(x_1) := & \min_{x_2, \dots, x_n} \sum_{i=2}^n (\alpha_{0i} + \beta'_{0i} x_1) x_i + \sum_{i=2}^n \beta_{0i} x_i^2 \\ \text{s.t.} & \sum_{i=2}^n (\alpha_{1i} + \beta'_{1i} x_1) x_i + \sum_{i=2}^n \beta_{1i} x_i^2 \leq \rho - \alpha_1 x_1 \end{aligned}$$

and $\min_{x_1} g_0(x_1)$ is equal to the optimal value of the original problem.

2.5 The case of composite separable constraints

We now consider the second special case of problem (NCP) in the Introduction:

$$\begin{aligned} \min \quad & h_0(x) + g_0(\beta_{01}f_1(x_1), \dots, \beta_{0n}f_n(x_n)) \\ \text{s.t.} \quad & h_l(x) + g_l(\beta_{l1}f_1(x_1), \dots, \beta_{ln}f_n(x_n)) \leq \rho_l \quad (l \in L), \end{aligned} \quad (46)$$

where $\rho_l \in \mathbb{R}$, $\beta_l = (\beta_{li})_{i=1}^n \in \mathbb{R}^n$ are given vectors, and $f = (f_1, \dots, f_n)$, $h_l, g_l : \mathbb{R}^n \rightarrow \mathbb{R}$ are given functions (for $l \in \{0\} \cup L$). We refer to problem (46) as the ‘composite separable’ case. We will assume the following:

Assumption 1. *The functions f, g_l, h_l ($l \in \{0\} \cup L$) are convex differentiable.*

Assumption 2. *For all $i \in \{1, \dots, n\}$ and $x \in \mathbb{R}^n$, either $\frac{\partial h_0}{\partial x_i}(x) > 0$ and $\frac{\partial h_l}{\partial x_i}(x) \geq 0 \quad \forall l \in L$, or $\frac{\partial h_0}{\partial x_i}(x) < 0$ and $\frac{\partial h_l}{\partial x_i}(x) \leq 0 \quad \forall l \in L$.*

Remark 2.4. *Under assumption 1, problem (46) is not necessarily convex. This is the case, for instance, when some of the β_{0i} ’s are negative. Even if $\beta_0 \geq 0$, the function $g(\beta_{01}f_1(x_1), \dots, \beta_{0n}f_n(x_n))$ is not necessarily convex since g is not assumed to be componentwise monotone increasing.*

Given scalars $a \in \mathbb{R}$ and $b, c_1, \dots, c_n > 0$, the function: $x \in \mathbb{R}^n \mapsto a + b \exp(\sum_{i=1}^n c_i x_i)$ is convex and its gradient is positive everywhere. Hence we can choose, for instance, for the functions h, h_l any positive linear combination of functions of this form. Another such an example is the function $x \in \mathbb{R}^n \mapsto (x_1 x_2 \dots x_n)^{1/n}$.

For the functions g, g_l , we can choose, for instance, any linear function, or the function: $x \in \mathbb{R}^n \mapsto \sum_{i=1}^n x_i^2$. In the latter case, the relaxation (47) becomes interesting only when the functions f_i are such that $\sum_{i=1}^n f_i^2$ is not convex (for, otherwise, the original problem (46) is convex already). This is the case for instance for $f_i(x_i) = -x_i^{c_i}$ where $0 < c_i < 1/2$, or $f_i(x_i) = -\sqrt{x_i} - c_i$ ($c_i > 0$).

A natural relaxation of (46) is obtained by linearizing the terms $f_i(x_i)$ by new variables y_i , leading to the following problem:

$$\begin{aligned} \min \quad & h_0(x) + g_0(\beta_{01}y_1, \dots, \beta_{0n}y_n) \\ \text{s.t.} \quad & h_l(x) + g_l(\beta_{l1}y_1, \dots, \beta_{ln}y_n) \leq \rho_l \quad (l \in L) \\ & y_i \geq f_i(x_i) \quad (i = 1, \dots, n) \end{aligned} \quad (47)$$

which, in view of assumption 1, is a convex problem. Clearly, (47) has a Slater point if (46) is strictly feasible.

We show next that problem (46) possesses a hidden convexity property.

Theorem 2.8. *Assume that assumptions 1-2 hold. Assume moreover that program (47) admits a Slater point and that it has an optimal solution. Then any optimal solution of (47) is an E-solution and thus (47) is equivalent to (46).*

Proof. Let (\bar{x}, \bar{y}) be an optimal solution of (47). As (47) has a Slater point, the KKT conditions hold: There exist scalars $\lambda_l \geq 0$ ($l \in L$) and $\mu_i \geq 0$ ($i = 1, \dots, n$) satisfying

$$\frac{\partial h_0}{\partial x_i}(\bar{x}) + \sum_{l \in L} \lambda_l \frac{\partial h_l}{\partial x_i}(\bar{x}) + \mu_i f'_i(\bar{x}_i) = 0 \quad \forall i = 1, \dots, n, \quad (48)$$

$$\beta_{0i} \frac{\partial g_0}{\partial y_i}(\beta_{01}\bar{y}_1, \dots, \beta_{0n}\bar{y}_n) + \sum_{l \in L} \lambda_l \beta_{li} \frac{\partial g_l}{\partial y_i}(\beta_{l1}\bar{y}_1, \dots, \beta_{ln}\bar{y}_n) - \mu_i = 0 \quad \forall i = 1, \dots, n \quad (49)$$

and $\mu_i = 0$ if $f_i(\bar{x}_i) < \bar{y}_i$. By assumption 2, we may assume that $\frac{\partial h_0}{\partial x_i}(\bar{x}) > 0$ and $\frac{\partial h_l}{\partial x_i}(\bar{x}) \geq 0$. Hence, if $f_i(\bar{x}_i) < \bar{y}_i$ then $\mu_i = 0$, and we get a contradiction in (48), since then the left hand side of (48) is strictly positive. Hence, an optimal solution (\bar{x}, \bar{y}) of (47) must have $f_i(\bar{x}_i) = \bar{y}_i$ for all i , i.e., it must be an E-solution. \square

The following corollary specializes this theorem to problem (5) with specific sign conditions.

Corollary 2.1. *Consider problem (5), where the functions f_1, \dots, f_n are convex and differentiable. Assume that problem (5) has a Slater point and an optimal solution. Suppose that the following sign conditions on the data α_l hold:*

$$\forall i \in \{1, \dots, n\} \text{ either } \alpha_{0i} > 0, \alpha_{li} \geq 0 \forall l \in L, \text{ or } \alpha_{0i} < 0, \alpha_{li} \leq 0 \forall l \in L, \quad (50)$$

Then every optimal solution is an E-solution and thus problems (3) and (5) are equivalent.

Proof. The proof immediately follows from Theorem 2.8, by taking $h_l(x) = \alpha_l^T x$ and $g_l(z) = \sum_{i=1}^n z_i$, for all $l \in \{0\} \cup L$, and noticing that condition (50) is required to validate assumption 2. \square

Extension to the case of block-separable variables

The result of Theorem 2.8 extends to the setting of block-variables, when the variables are partitioned into groups of variables $x_{[1]}, \dots, x_{[N]}$. We now assume that f_j ($j = 1, \dots, N$) is a convex function from $\mathbb{R}^{|[j]|}$ to \mathbb{R} and that g, g_l are convex functions from \mathbb{R}^N to \mathbb{R} , and we consider the following analogue of problem (46):

$$\begin{aligned} \min \quad & h_0(x) + g_0(\beta_{01}f_1(x_{[1]}), \dots, \beta_{0N}f_N(x_{[N]})) \\ \text{s.t.} \quad & h_l(x) + g_l(\beta_{l1}f_1(x_{[1]}), \dots, \beta_{lN}f_N(x_{[N]})) \leq \rho_l \quad (l \in L) \end{aligned} \quad (51)$$

and its relaxation:

$$\begin{aligned} \min \quad & h_0(x) + g_0(\beta_{01}y_1, \dots, \beta_{0N}y_N) \\ \text{s.t.} \quad & h_l(x) + g_l(\beta_{l1}y_1, \dots, \beta_{lN}y_N) \leq \rho_l \quad (l \in L) \\ & f_j(x_{[j]}) \leq y_j \quad (j = 1, \dots, N). \end{aligned} \quad (52)$$

Consider the following analogue of assumption 2:

Assumption 3. *For all $j \in \{1, \dots, N\}$, there exists $i \in [j]$ such that, for all $x \in \mathbb{R}^n$, either $\frac{\partial h_0}{\partial x_i}(\bar{x}) > 0$ and $\frac{\partial h_l}{\partial x_i}(\bar{x}) \geq 0 \forall l \in L$, or $\frac{\partial h_0}{\partial x_i}(\bar{x}) < 0$ and $\frac{\partial h_l}{\partial x_i}(\bar{x}) \leq 0 \forall l \in L$.*

Theorem 2.9. *Consider problem (51), where the functions f_j , $j = 1, \dots, N$, and h_l, g_l , $l \in \{0\} \cup L$, are convex and assume that assumption 3 holds. If problem (52) has a Slater point and attains its minimum, then every optimal solution of (52) is an E-solution and thus problems (51) and (52) are equivalent.*

Proof. Use the KKT conditions. \square

3 Application to Robust Optimization

The Robust Optimization (RO) methodology addresses optimization problems affected by parameter uncertainty in constraints. We focus here on a linear constraint

$$(RC) \quad a^T u + b^T v \geq \eta, \quad (53)$$

where $u \in \mathbb{R}^n$ and $v \in \mathbb{R}^p$ are the decision variables, $b \in \mathbb{R}^p$ is a fixed vector and η a fixed scalar, and $a \in \mathbb{R}^n$ is an uncertain vector, which is only known to reside in a set \mathcal{A} (called the *uncertainty set*). The robust counterpart of (53) is then

$$a^T u + b^T v \geq \eta, \quad \forall a \in \mathcal{A}. \quad (54)$$

With this formulation, it is required to determine the values of both u and v before the actual realization of a is available (“here and now” decisions). However, in many important applications, notoriously in multi-period (dynamic) decision problems, only the variable u has to be fixed a priori (a nonadjustable variable) while v can be determined after a (or a part of it) is revealed. In such situations the relevant RO model is the so-called Adjustable RO (ARO), where instead of (54), the interpretation of (RC) is

$$(ARC) \quad \begin{cases} \exists u \forall a \in \mathcal{A} \exists v = v(a) \text{ such that } u, v \text{ satisfy} \\ a^T u + b^T v \geq \eta. \end{cases} \quad (55)$$

Solving (ARC) is in most cases an NP-hard problem. The Affinely Adjustable Robust Counterpart (AARC) method thus restrict the function $v(a)$ to be *affine*:

$$v(a) = v^0 + \sum_{i=1}^n a_i v^i, \quad (56)$$

with $v^i \in \mathbb{R}^p$, for $i = 0, 1, \dots, n$. The expression (56) is termed a Linear Decision Rule (LDR). The corresponding interpretation of (53) then becomes

$$(AARC) \quad a^T u + b^T \left(v^0 + \sum_{i=1}^n a_i v^i \right) \geq \eta \quad \forall a \in \mathcal{A}, \quad (57)$$

or equivalently, a single inequality

$$\min_{a \in \mathcal{A}} \left\{ a^T u + b^T \left(v^0 + \sum_{i=1}^n a_i v^i \right) \right\} \geq \eta \quad (58)$$

in the variables $u \in \mathbb{R}^n, v^i \in \mathbb{R}^p (i = 0, \dots, n)$.

Problem (58) is tractable for a large variety of uncertainty sets \mathcal{A} . The “price” to get this tractability was the restriction of the adjustable variable $v(a)$ to be an LDR.

Consider the case where \mathcal{A} is described by a system of separable-type inequalities:

$$\mathcal{A}_s = \{ a \in \mathbb{R}^n \mid \alpha_l^T a + \beta_l^T f(a) \leq \rho_l, \quad l \in L \}, \quad (59)$$

where $f_i : \mathbb{R} \mapsto \mathbb{R}$ are given univariate functions, and $f = (f_1, \dots, f_n)$. In this case the hidden convexity results in the previous sections offer an opportunity to use more general functions $v(a)$ than an affine one. In particular let $v(a)$ be given as a Separable Decision Rule (SDR) of the form

$$v(a) = v^0 + \sum_{i=1}^n v^i a_i + \sum_{i=1}^n w^i f_i(a_i), \quad (60)$$

where $v^0, v^i, w^i \in \mathbb{R}^p$, for $i = 1, \dots, n$. We emphasize that in this SDR the same separable functions are chosen as those given in the constraint functions that define the uncertainty region. The corresponding interpretation of (53) then becomes

$$(SARC) \quad b^T v^0 + \min_{a \in \mathcal{A}_s} \left\{ \sum_{i=1}^n a_i (u_i + b^T v^i) + \sum_{i=1}^n b^T w^i f_i(a_i) \right\} \geq \eta.$$

Collecting all variables u_j, v_j^0, v_j^i, w_j^i ($\forall i, \forall j$) in a single vector x , problem (SARC) reduces to the following generic problem: find x such that

$$\gamma(x) + \min_{a \in \mathcal{A}_s} \{ \alpha_0(x)^T a + \beta_0(x)^T f(a) \quad : \quad \alpha_l^T a + \beta_l^T f(a) \leq \rho_l \quad (l \in L) \} \geq \eta, \quad (61)$$

where the functions $\alpha_0(x)$, $\beta_0(x)$ and $\gamma(x)$ are affine in x .

The following example illustrates this reformulation.

Example 3.1. Consider the following robust counterpart:

$$(1 + a_1)u_1 + a_2u_2 + v \geq 1 \quad \forall (a_1, a_2) \quad : \quad a_1^2 + 2a_2^2 \leq 2, \quad (62)$$

where v is adjustable and for which we use the following SDR (of the form (60)):

$$v = v^0 + v^1 a_1 + v^2 a_2 + w^1 a_1^2 + w^2 a_2^2.$$

Substitution of this SDR into (62) yields:

$$v^0 + u_1 + a_1(u_1 + v^1) + a_2(u_2 + v^2) + w^1 a_1^2 + w^2 a_2^2 \geq 1 \quad \forall (a_1, a_2) \quad : \quad a_1^2 + 2a_2^2 \leq 2. \quad (63)$$

By defining $f_1(a_1) = a_1^2$, $f_2(a_2) = a_2^2$, and $x = (u_1, u_2, v^0, v^1, v^2, w^1, w^2)^T$, $\gamma(x) = x_1 + x_3$, $\alpha_{01}(x) = x_1 + x_4$, $\alpha_{02}(x) = x_2 + x_5$, $\beta_{01}(x) = x_6$, $\beta_{02}(x) = x_7$, (62) can be equivalently written into the format of (61):

$$\gamma(x) + \min_{(a_1, a_2)} \{ \alpha_{01}(x)a_1 + \alpha_{02}(x)a_2 + \beta_{01}(x)f_1(a_1) + \beta_{02}(x)f_2(a_2) \quad : \quad f_1(a_1) + 2f_2(a_2) \leq 2 \} \geq 1.$$

The minimization problem in the left hand side of (61) is of the same format as problem (3) studied in Section 2. Using the hidden convexity results there, the following result is obtained.

Theorem 3.1. Assume that f_1, \dots, f_n are convex, $\beta_l \geq 0$ for all $l \in L$, and that, for each $i \in \{1, \dots, n\}$, $\beta_{li} \neq 0$ for some $l \in L$. Assume moreover that the uncertainty region \mathcal{A}_s (59) has a strictly feasible solution, and that the m constraints that define it are termwise parallel. Then x is an SARC solution (i.e., satisfies inequality (61)) if and only if there exists $v \in \mathbb{R}^m$ such that (x, v) is a feasible solution for the following convex inequalities:

$$\begin{cases} \sum_{i=1}^n (\beta_{0i}(x) + \sum_{l \in L} \beta_{li} v_l) f_i^* \left(-\frac{\alpha_{0i}(x) + \sum_{l \in L} \alpha_{li} v_l}{\beta_{0i}(x) + \sum_{l \in L} \beta_{li} v_l} \right) + \sum_{l \in L} \rho_l v_l + \eta \leq 0 \\ \beta_{0i}(x) + \sum_{l \in L} \beta_{li} v_l \geq 0 \quad (i = 1, \dots, n) \\ v \geq 0. \end{cases} \quad (64)$$

Proof. The proof follows immediately from Theorems 2.1, 2.2 and 2.5. □

Example 3.2. (Continuation of Example 3.1.) Using Theorem 3.1 and observing that $f_i^*(s) = \frac{1}{4}s^2$ for $i = 1, 2$, we obtain the following result for the Robust Counterpart (63):

$$\begin{cases} \frac{(x_1 + x_4)^2}{4(x_6 + v)} + \frac{(x_2 + x_5)^2}{4(x_7 + 2v)} + 2v + 1 \leq 0 \\ x_6 + v \geq 0 \\ x_7 + 2v \geq 0 \\ v \geq 0. \end{cases} \quad (65)$$

Note that this system of inequalities can be rewritten as conic quadratic inequalities.

Example 3.3. In [4] it is proposed, in case the uncertain parameter a in (3) can be considered as a probability vector, to use uncertainty regions defined by so-called ϕ -divergence functions. Let us concentrate in this example on the well-known Kullback-Leibler function:

$$f_i(a_i) = a_i \log \left(\frac{a_i}{a_i^0} \right),$$

where a_i^0 ($i = 1, \dots, n$) are given nominal values. In this case the uncertainty region is defined by

$$\mathcal{A}_s = \left\{ a \in \mathbb{R}^n \mid \sum_{i=1}^n a_i \log \left(\frac{a_i}{a_i^0} \right) \leq \rho \right\}.$$

If we use the following separable decision rule:

$$v(a) = v^0 + \sum_{i=1}^n a_i v^i + \sum_{i=1}^n w^i a_i \log \left(\frac{a_i}{a_i^0} \right), \quad (66)$$

then by Theorem 3.1 we obtain that the corresponding (SARC) is given by (64), where $f_i^*(s) = a_i^0 e^{s-1}$.

Remark 3.1. Theorem 3.1 can be easily generalized for the block-separable case.

4 Complexity results

In this section we analyze the complexity of the problems studied in the previous sections. A key notion in complexity theory for convex optimization problems is *self-concordance*. In [7] it is shown that if the optimization problem admits a self-concordant barrier, then there are interior point methods that yields an ϵ -optimal solution in $O(\sqrt{n} |\ln \epsilon|)$ Newton steps, where n is the number of variables. In this section we show that for many cases the logarithmic barrier function for the optimization problems studied in the previous sections are self-concordant, thereby proving that there are interior point methods that solve such problems in polynomial time.

We first start with the formal definition of the logarithmic barrier function and self-concordance.

Definition 4.1. The logarithmic barrier function for the set of constraints

$$\{g_i(z) \leq 0, \quad i \in I\}$$

in the variables $z \in \mathbb{R}^n$, is defined as

$$\varphi(z) = - \sum_{i \in I} \ln(-g_i(z)).$$

□

Definition 4.2. Let $F \subset \mathbb{R}^n$ be an open and convex set and κ a nonnegative number. A function $\varphi : F \rightarrow \mathbb{R}$ is called κ -self-concordant on F if φ is convex, belongs to $C^3(F)$, and satisfies:

$$|\nabla^3 \varphi(y)[h, h, h]| \leq 2\kappa (h^T \nabla^2 \varphi(y) h)^{\frac{3}{2}}, \quad \forall y \in F \quad \forall h \in \mathbb{R}^n,$$

where $\nabla^3 \varphi(y)[h, h, h]$ denotes the third order differential of φ at y and h . We call φ self-concordant if it is κ -self-concordant for some κ . □

The primal relaxation problem (5) for all the cases studied in Section 2 have linear objectives and linear constraints and additional nonlinear convex constraints

$$f_i(x_i) \leq y_i, \quad i = 1, \dots, n. \quad (67)$$

Note that the logarithmic barrier function for the linear constraints is self-concordant, hence we only have to analyze the logarithmic barrier function for the convex constraints (67), i.e.,

$$\varphi(x, y) = - \sum_{i=1}^n \ln(y_i - f_i(x_i)).$$

For several choices of f_i we may be able to prove that the corresponding logarithmic barrier is self-concordant by using the following lemma taken from [6].

Lemma 4.1. *Assume that a convex function $f \in C^3(\mathbb{R}^+)$ satisfies $|f'''(s)| \leq \kappa f''(s)/s$ for some $\kappa > 0$. Then the logarithmic barrier function for the constraints*

$$\{f(s) \leq t, \quad s \geq 0\},$$

i.e., $-\ln(t - f(s)) - \ln s$, is $(1 + \frac{1}{3})\kappa$ -self-concordant. □

Hence, if all f_i satisfy the condition in the lemma, then the logarithmic barrier function for (5) is self-concordant, and hence there are interior point methods that solve these problems in polynomial time. It is easy to see that e.g. $f_i(x_i) = x_i^2$ or $f_i(x_i) = -\ln x_i$ satisfies the condition of the lemma. For other choices we may first need a reformulation of the constraint. Suppose e.g. that $f_i(x_i) = e^{x_i}$. Then the constraint $f_i(x_i) \leq y_i$ is equivalent to $-\ln y_i \leq -x_i$. Now $-\ln y_i$ satisfies the conditions of the lemma, and hence the corresponding logarithmic barrier function is self-concordant. In Table 2 several examples are given that admit self-concordant barrier functions for the primal problem.

Since in many problems studied in this paper the number of constraints is low, working with the dual may improve the complexity. Let us first consider problem (8) where $m = 1$. This is a 1 variable optimization problem. By using e.g. binary search or golden section search we obtain an ϵ -accurate solution in $O(|\ln \epsilon|)$ iterations. Each iteration costs $O(n)$ number of basic arithmetic operations. Hence, the overall complexity to solve problem (8), and thus problem (3), is $O(n|\ln \epsilon|)$.

All dual problems (8) and (64) have the same structure and their constraints can be reformulated as follows:

$$\begin{cases} \sum_{i=1}^n t_i \leq \sigma_0(z) \\ r_i f_i^*(s_i/r_i) \leq t_i, \quad i = 1, \dots, n \\ r_i = \sigma_i(z), \quad i = 1, \dots, n \\ s_i = \tau_i(z), \quad i = 1, \dots, n \\ r_i \geq 0, \quad i = 1, \dots, n, \end{cases} \quad (68)$$

in which σ_i , $i = 0, \dots, n$, and τ_i , $i = 1, \dots, n$, are linear functions in the optimization variables z .

We illustrate this reformulation for problem (8):

$$\begin{cases} z^T = (v^T, w) \\ \sigma_0(z) = - \sum_{l \in L} \rho_l v_l + w \\ \sigma_i(z) = \beta_{0i} + \sum_{l \in L} \beta_{li} v_l, \quad i = 1, \dots, n \\ \tau_i(z) = \alpha_{0i} + \sum_{l \in L} \alpha_{li} v_l, \quad i = 1, \dots, n. \end{cases} \quad (69)$$

For several choices of f_i we may be able to prove that there exists a self-concordant logarithmic barrier for (68) by using the following lemma taken from [4]:

$f(t)$	primal problem	$f^*(s)$ (domain)	dual problem
t	LP	0 ($s = 1$)	LP
t^2	QP	$s^2/4$ ($s \in \mathbb{R}$)	CQP
$ t ^p/p$ ($p > 1$)	CQP	$ s ^q/q$ ($s \in \mathbb{R}$)	CQP
$-t^p/p$ ($t \geq 0, 0 < p < 1$)	CQP	$-(-s)^q/q$ ($s \leq 0$)	CQP
$-\log t$ ($t > 0$)	S.C.	$-1 - \log(-s)$ ($s < 0$)	S.C.
$t \log t$ ($t > 0$)	S.C.	e^{s-1} ($s \in \mathbb{R}$)	S.C.
e^t	S.C.	$\begin{cases} s \log s - s & (s > 0) \\ 0 & (s = 0) \end{cases}$	S.C.
$\log(1 + e^t)$	S.C.	$\begin{cases} s \log s + (1 - s) \log(1 - s) & (0 < s < 1) \\ 0 & (s = 0, 1) \end{cases}$	S.C.
$\sqrt{1 + t^2}$	CQP	$-\sqrt{1 - s^2}$ ($-1 \leq s \leq 1$)	CQP

Table 2: Some examples for f , with their conjugates, and indications of the tractability of the primal and dual problems; S.C. means “admits self-concordant barrier”, (C)QP means (Conic) Quadratic Programming, LP means Linear Programming. The parameters p and q are related as follows: $1/p + 1/q = 1$.

Lemma 4.2. *Assume that a convex function $f \in C^3(\mathbb{R}^+)$ satisfies $|f'''(s)| \leq \kappa f''(s)/s$ for some $\kappa > 0$. Then the logarithmic barrier function for the constraints*

$$\{rf(s/r) \leq t, r \geq 0, s \geq 0\},$$

i.e., $-\ln(t - rf(s/r)) - \ln r - \ln s$, is $(2 + \frac{\sqrt{2}}{3})\kappa$ -self-concordant. □

Hence, if all f_i^* in (68) satisfy the condition in the lemma, then the logarithmic barrier function for (68) is self-concordant, and hence interior point methods solve these problems in polynomial time. Some examples for f_i^* that satisfy the condition of the above lemma are x_i^2 and $-\ln x_i$. In Table 2 several examples are given that admit self-concordant barrier functions for the dual problem.

We conclude this section by the remark that the tractability results for both the primal and dual problems given in Table 2 also hold after affine transformations of the variable and for summations of functions stated in Table 2. Let us first start with the primal problem (5). It is well-known that a function that is self-concordant remains self-concordant after addition, scaling, and affine transformations of its variables.

Before we give an example of the effect of affine transformations, addition and scaling on the tractability of the dual problem, we first state a well-known result for the conjugate of the sum of functions.

Proposition 4.1. [9] *Assume that $g = \sum_k g_k$, and g_k are convex, and the intersection of the relative interiors of the domains of g_k is nonempty, i.e., $\cap_k \text{ri}(\text{dom } g_k) \neq \emptyset$. Then*

$$\left(\sum_k g_k \right)^*(s) = \inf_{\sum_k s_k = s} \sum_k g_k^*(s_k),$$

and the infimum is attained for some s_k . □

Suppose now that $f_i = \sum_k f_{ik}$ and $\cap_k \text{ri}(\text{dom } f_{ik}) \neq \emptyset$, $i = 1, \dots, n$, then (68) is equivalent to

$$\begin{cases} \sum_{i,k} t_{ik} \leq \sigma_0(z) \\ y_i f_{ik}^*(s_{ik}/y_i) \leq t_{ik}, \quad i = 1, \dots, n; \quad \forall k \\ y_i = \sigma_i(z), \quad i = 1, \dots, n \\ \sum_k s_{ik} = \tau_i(z), \quad i = 1, \dots, n \\ y_i \geq 0, \quad i = 1, \dots, n. \end{cases} \quad (70)$$

Example for dual problem. Suppose $f_i(x_i) = 3x_i \log x_i + e^{x_i-1}$. Then f_i^* cannot be stated in explicit form. However, since $f_i = f_{i1} + f_{i2}$, where $f_{i1}(x_i) = 3x_i \log x_i$, and $f_{i2}(x_i) = e^{x_i-1}$, we can use (70) since f_{i1}^* and f_{i2}^* can be stated in explicit form.

References

- [1] A. Ben-Tal, M. Teboulle. Hidden convexity in some nonconvex quadratically constrained quadratic programming. *Mathematical Programming*, 72, 51–63, 1995.
- [2] A. Ben-Tal, L. El-Ghaoui, A. Nemirovski. *Robust Optimization*. Princeton Series in Applied Mathematics, 2009.
- [3] A. Ben-Tal, D. den Hertog. Immunizing conic quadratic optimization problems against implementation errors. CentER Discussion Paper, Tilburg University, 2011.
- [4] A. Ben-Tal, D. den Hertog, A. De Waegenaere, B. Melenberg, G. Rennen. Robust solutions of optimization problems affected by uncertain probabilities. CentER Discussion Paper, Tilburg University, 2011.
- [5] A. Ben-Tal, A. Goryashko, E. Guslitzer, A. Nemirovski. Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, 99, 351–376, 2004.
- [6] D. den Hertog. *Interior Point Approach to Linear, Quadratic and Convex Programming*. Kluwer Academic Publishers, Dordrecht, 1994.
- [7] Y.E. Nesterov and A. Nemirovsky. *Interior Point Polynomial Algorithms in Convex Programming*. SIAM Studies in Applied Mathematics, Philadelphia, 1993.
- [8] I. Polik, T. Terlaky. A survey of the S-lemma. *SIAM Review*, 49(3), 371–418, 2007.
- [9] R.T. Rockafellar. *Convex Analysis*. Princeton Press, Princeton, 1970.
- [10] R.J. Stern, H. Wolkowicz. Indefinite trust region subproblems and nonsymmetric eigenvalue perturbations. *SIAM Journal on Optimization*, 5, 286–313, 1995.
- [11] F. Uhlig. Definite and semidefinite matrices in a real symmetric matrix pencil. *Pacific Journal of Mathematics*, 49 (2), 561–568, 1972.
- [12] V.A. Yakubovich. Minimization of quadratic functionals under quadratic constraints and the necessity of a frequency condition in the quadratic criterion for absolute stability of nonlinear control systems. *Soviet Math. Dokl.*, 14, pp. 593-597, 1973.
- [13] Y. Ye, S. Zhang. New results on quadratic minimization. *SIAM Journal on Optimization*, 14 (1), 245–267, 2003.