Computation of invariant sets for complex systems

Benoît Legat Raphaël M. Jungers

January 7, 2023

Contents

1 Introduction

In the analysis and control of complex systems such as hybrid or switched systems, the computation of sets satisfying given properties is paramount. Let us mention for instance safety analysis, model predictive control and Information-Theoretic bounds for communication networks. Mathematically, the computation of such sets is best handled by the formulation of a corresponding optimization problem.

In numerical optimization, the introduction of a modeling layer on top of optimization solvers has been a key enabler, allowing different communities to share emergent technologies more easily. When the purpose is to compute sets rather than numbers (a.k.a. 'set programming'), the numerical complexity is amplified, yet there does not currently exist such a clearly defined interface. Even the meaning of 'computing a set' is not clear and actually depends on the intended end-use.

The purpose of this chapter is first to unify many control-theoretic problems under the umbrella of set programming. Then, we survey diverse numerical and theoretic techniques, that constitute a rich toolbox for handling such problems. We finally point out some current challenges. In section 2, we show a list of applications where the problem can be encoded in a set program. In section 3, we provide a unifying view on these problems by representing the model at the level of abstract sets without using any notion specific to a certain *template* (i.e., family of sets). We develop several results that are useful for both theory and computation, relying on fixed point properties and convex analysis tools such as duality, polarity or support functions. In section 4, we particularize the generic results of section 3 to polyhedra, zonotopes, ellipsoids, piecewise semi-ellipsoids, and semialgebraic sets. We discuss their computational and numerical properties. In section 5, we discuss different approaches for providing approximation guarantees when restricting a set program to a given template.

The results presented rely on convex analysis prerequisites that can be found in [46] as well as polyhedral computation results that can be found in [50]. We provide such background material in appendix.

1.1 Notation

The set $\{1, \dots, m\}$ is denoted $[m]$. Given a subset $I \subseteq [n]$, $\pi_{I,n}$ denotes the projection matrix on the coordinates I. Given a linear subspace V, π_V denotes an orthogonal projection matrix with range V. We denote the n-dimensional nonnegative orthant as $\mathbb{R}^n_+$ and the $n-1$-*dimensional standard simplex* as $\Delta^{n-1} = \{\, \lambda \in \mathbb{R}^n_+ \mid \lambda_1 + \cdots + \lambda_n = 1 \,\}$. We denote the subset of $n \times n$ symmetric matrices as $\mathcal{S}^n \subseteq \mathbb{R}^{n\times n}$, its subset of positive semidefinite matrices as $\mathcal{S}^n_+ \subseteq \mathcal{S}^n$ and its subset of positive definite matrices as $\mathcal{S}^n_{++} \subseteq \mathcal{S}^n_+$. Given a matrix $A \in \mathcal{S}^n_+$, its pseudoinverse is denoted $A^\dagger$.

Given a set $\mathcal{S}$, $\operatorname{aff}(\mathcal{S})$ denotes its affine hull, $\operatorname{conv}(\mathcal{S})$ denotes its convex hull, $\operatorname{cone}(\mathcal{S})$ denotes the convex cone generated by $\mathcal{S}$ and $\mathcal{S}^\circ$ denotes its polar, defined in definition A.6.

Consider a vector-valued function $f : \mathbb{R}^n \to \mathbb{R}^r$. Given a set $\mathcal{S} \subseteq \mathbb{R}^n$, its image under the function f, denoted by $f(\mathcal{S})$, is the set $\{\, f(x) \mid x \in \mathcal{S} \,\}$. Given a set $\mathcal{S} \subseteq \mathbb{R}^r$, its preimage under the function f, denoted by $f^{-1}(\mathcal{S})$, is the set $\{\, x \in \mathbb{R}^n \mid f(x) \in \mathcal{S} \,\}$.

2 Set programming formulations for the control of complex systems

In this chapter, we introduce optimization programs for which decision variables are sets and their applications.

2.1 Numerical implementation

The theoretical results of this chapter provide the foundation for the SetProg software. This Julia package extends JuMP [17] to allow modeling set decision variables and inclusion or membership constraints with these variables. As summarized in fig. 2, SetProg first determines whether the primal or dual representation should be used for each set and then formulates the optimization program as a linear, semidefinite or Sum-of-Squares program depending on the template chosen for each set. SumOfSquares [49] then reformulates the

Sum-of-Squares constraint into semidefinite constraints and communicates the resulting program to the solver through the solver-independent MathOptInterface [32]. As a result, SetProg provides a seemless integration of set programming into a generic optimization framework as illustrated in the following example.

Example 2.1. Consider the following matrices inspired from [2, Example 5.4].

$$4A_1 = \begin{pmatrix} -1 & -1 \\ -4 & 0 \end{pmatrix}, \qquad 4A_2 = \begin{pmatrix} 3 & 3 \\ -2 & 2 \end{pmatrix}. \tag{1}$$

We consider the following set programs where the L_1 objective is detailed section 4. The SetProg syntax is displayed on the right and the optimal solutions for different templates are given in fig. 1. The variables `A1` are `A2` are the matrices A_1, A_2 defined in eq. (1), `solver` is any solver implementing the MathOptInterface (the solver used to generate the solutions displayed in fig. 1 is CSDP [14]) and `template` depends on the chosen template.

$$\begin{aligned} \underset{\mathcal{S}}{\text{minimize}} \quad & L_1(\mathcal{S}, [-1,1]^2) \\ \text{subject to:} \quad & \mathcal{S} \supseteq [-1,1]^2 \\ & A_1\mathcal{S} \subseteq \mathcal{S} \\ & A_2\mathcal{S} \subseteq \mathcal{S}. \end{aligned} \tag{2}$$

$$\begin{aligned} \underset{\mathcal{S}}{\text{maximize}} \quad & L_1(\mathcal{S}, [-1,1]^2) \\ \text{subject to:} \quad & \mathcal{S} \subseteq [-1,1]^2 \\ & A_1\mathcal{S} \subseteq \mathcal{S} \\ & A_2\mathcal{S} \subseteq \mathcal{S} \\ & \mathcal{S} \text{ is convex.} \end{aligned} \tag{3}$$

```
□ = HalfSpace([1, 0], 1) ∩ HalfSpace([-1, 0], 1) ∩
     HalfSpace([0, 1], 1) ∩ HalfSpace([0, -1], 1)
model = Model(solver)
@variable(model, S, template)
@objective(model, Min, L1_heuristic(volume(S), □))
@constraint(model, S ⊇ □)
@constraint(model, A1 * S ⊆ S)
@constraint(model, A2 * S ⊆ S)
```

The numerical results of this chapter can be reproduced using the associated Code Ocean capsule [27].

2.2 Set programs

It is important to distinguish conic programs, for which the decision variables are *vectors* that belongs to *given cones*, from set programs where the decision variables are *sets* that belongs to *given families* of sets, called *templates*.

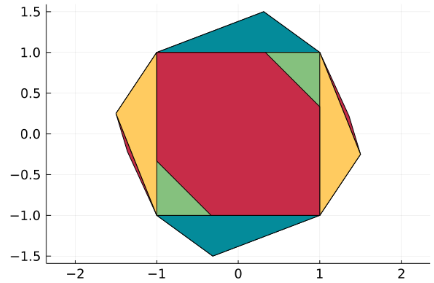
This distinction is similar to the extension from first-order logic to second-order logic. In first-order logic, propositions quantify only over variables that represent elements of given sets while in second-order logic, propositions can quantify over variables that may represent sets that are elements of given families of sets. When considering sets inside propositions, Russell exhibited the famous Russell's paradox [47]: "Let $\mathcal{S} = \{\, x \mid x \notin x \,\}$, then $\mathcal{S} \in \mathcal{S} \Leftrightarrow \mathcal{S} \notin \mathcal{S}$". This paradox is resolved in second-order logic by using a hierarchy of sets that can only contain sets of the lower order. For instance, in second-order logic, the variables are sets that cannot contain sets themselves. The same applies for set programming, the decision variables represent subsets of $\mathbb{R}^n$ or even convex subsets of $\mathbb{R}^n$ hence we shall not encounter Russell's paradox.

A key difference between the set programs that we study in this section and the optimization of *submodular functions* [36] is that the latter focuses on discrete sets while we focus on subsets of $\mathbb{R}^n$. Given a finite set Ω and a set function $h : 2^\Omega \to \mathbb{R}$, the function f is *submodular* if the condition holds for all subsets $J, K \subseteq \Omega$:

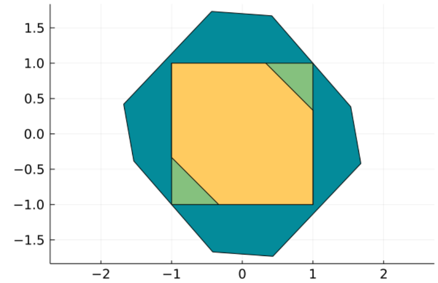
$$h(J) + h(K) \geq h(J \cup K) + h(J \cap K). \tag{4}$$

The concept of *submodular continuous functions* was introduced recently but the functions are defined for continuous vectors, not subsets of a continuous domain. Suppose $\Omega = [n]$, then, given a submodular function $h : 2^\Omega \to \mathbb{R}$, we can define the function $f : \{0,1\}^{|\Omega|} \to \mathbb{R}$ with $f(x) = h(\{\, \omega \in \Omega \mid x_\omega = 1 \,\})$. The submodularity condition (4) is then rewritten as (see e.g. [7, eq. (1)])

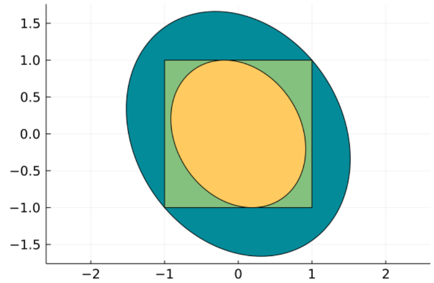
$$f(x) + f(y) \geq f(\min(x,y)) + f(\max(x,y))$$



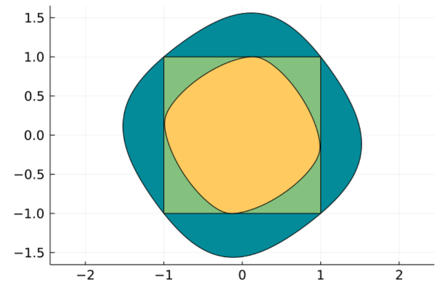
(a) Subsequent iterations of the fixed point algorithm; see section 4.1.1



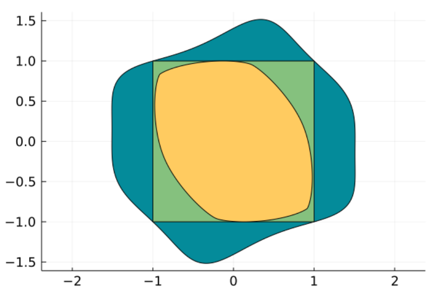
(b) Polytope template with face fan fixed to the face fan of an octogon; see section 4.1.2



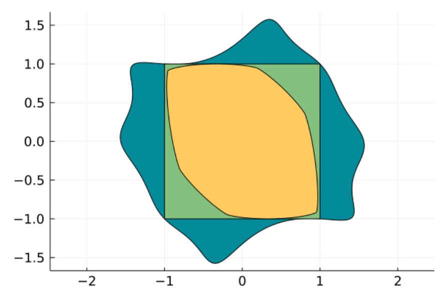
(c) Ellipsoidal template



(d) Polyset template of degree 4



(e) Polyset template of degree 6



(f) Polyset template of degree 8

Figure 1: Optimal solutions to the set programs of example 2.1 for different templates. The green set is $[-1,1]^2$. The blue set is the solution to set program (2) and the yellow set is the solution to set program (3).

where min and max are applied component-wise. This condition is used to define submodular continuous functions $f : [0,1]^{|\Omega|} \to \mathbb{R}$. If f is twice differentiable, it is equivalent to all off-diagonal entries of the hessian being nonpositives, see e.g. [9, eq. (2)].

The set program searching over all sets, not specific to any template, is called the *generic set program*. The generic set program restricted to sets of a given template T is called the *set program for T*. A generic set program is defined as follows:

$$\begin{aligned} \underset{\mathcal{S}_1 \subseteq \mathbb{R}^{n_1}, \ldots, \mathcal{S}_N \subseteq \mathbb{R}^{n_N}}{\text{minimize}} \quad & f(\mathcal{S}_1, \ldots, \mathcal{S}_N) \\ \text{subject to:} \quad & g_i(\mathcal{S}_1, \ldots, \mathcal{S}_N) \subseteq h_i(\mathcal{S}_1, \ldots, \mathcal{S}_N), \quad i = 1, 2, \ldots, M. \end{aligned} \tag{5}$$

where $\mathcal{S}_i \subseteq \mathbb{R}^{n_i}$ is a set decision variable, $f(\mathcal{S}_1, \ldots, \mathcal{S}_N)$ is the objective function and the inclusion constraints are given by $g_i(\mathcal{S}_1, \ldots, \mathcal{S}_N) \subseteq h_i(\mathcal{S}_1, \ldots, \mathcal{S}_N)$ for $i = 1, \ldots, M$. Note that this form also encapsulates membership constraints of the form $x \in \mathcal{S}$ as it can be encoded as an inclusion constraint $\{x\} \subseteq \mathcal{S}$.

In this chapter, we take the following approach to set programs, which is illustrated in fig. 2:

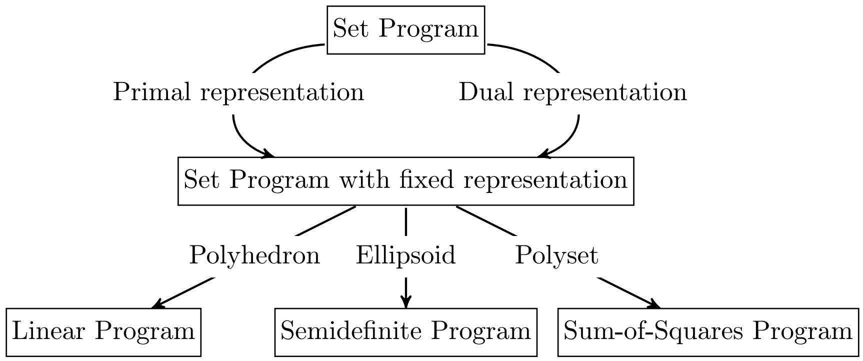


Figure 2: Illustration of the advantages of the set programming interface. First, the choice between primal representation (e.g. gauge function) or dual representation (e.g. support function) is done in a template-independent manner. Second, once the representation is determined for each set, each inclusion constraint can be reformulated independently of the rest of the set program.

1. First, given the properties of *gauge functions* (defined in definition A.2) and *support functions* (defined in definition A.5) and the program constraints, we determine whether each set variable should be represented with its gauge or support function.
2. Second, we consider the different templates and analyze the program obtained by formulating the program in terms of its gauge or support function, depending on what was determined in the previous section.

The advantage of this approach is that the first step is independent of the actual template and hence it gives a generic geometric approach to the computation of sets that are solutions to the set program instead of an algebraic template-dependent approach as discussed in section 3.

For polytopes, ellipsoids, piecewise semi-ellipsoids (resp. polysets and piecewise polysets), specific classes of set programs can be reformulated as a semidefinite program (resp. Sum-of-Squares program). The reformulation is done in the second step and, interestingly, the interdependence of the different constraints and the objective only appears in the choice of the representation, which is already carried out in a template-independent fashion in the first step. Therefore, the second-step can be done independently for the objective function and for each constraint. For this reason, we consider in each section, the constraints and the objective functions in isolation. Moreover, since each constraint can be reformulated independently, implementing only a few constraint reformulations enables the reformulation of programs made of any of their combinations, as long as there is a valid choice of representation.

2.3 Applications

In this section, we show how to formulate the computation of the invariant sets of discrete-time dynamical systems as set programs. To this end, we reformulate the invariance condition as a set inclusion. The remaining of the set program depends for instance on whether the computed set should have maximal (resp. minimal) volume while being included in (resp. including) some given set of constraints, etc...

A discrete-time system is defined by the iteration

$$x_{k+1} = f(x_k), \qquad x_k \in \mathcal{X} \tag{6}$$

where $f : \mathbb{R}^n \to \mathbb{R}^n$ and $\mathcal{X} \subseteq \mathbb{R}^n$.

Definition 2.1 (Invariant set)**.** A set $\mathcal{S} \subseteq \mathcal{X}$ is *invariant* for system (6) if for any state $x_0 \in \mathcal{S}$, the trajectory of the system with initial state x_0 remains in $\mathcal{S}$.

Proposition 2.1. A set $\mathcal{S} \subseteq \mathcal{X}$ is invariant for system (6), as defined in definition 2.1, if and only if

$$f(\mathcal{S}) \subseteq \mathcal{S}.$$

To represent uncertainty due to the imperfection of the model or actual noise in the system being modeled, a disturbance is added to the system:

$$x_{k+1} = f(x_k, w_k), \qquad x_k \in \mathcal{X}, w_k \in \mathcal{W} \tag{7}$$

where $f : \mathbb{R}^{n_x+n_w} \to \mathbb{R}^{n_x}$, $\mathcal{X} \subseteq \mathbb{R}^{n_x}$ and $\mathcal{W} \subseteq \mathbb{R}^{n_w}$.

While the disturbance is an external input of the system that cannot be chosen, the input can sometimes be controlled. This is modeled using a control input u:

$$x_{k+1} = f(x_k, u_k), \qquad x_k \in \mathcal{X}, u_k \in \mathcal{U} \tag{8}$$

where $f : \mathbb{R}^{n_x+n_u} \to \mathbb{R}^{n_x}$, $\mathcal{X} \subseteq \mathbb{R}^{n_x}$ and $\mathcal{U} \subseteq \mathbb{R}^{n_u}$.

With a disturbance or control input, the system is no longer autonomous. Each new iterate x_{k+1} depends not only on the previous iterate x_k but also on the external inputs w or u. In *descriptor systems*, there is no external input but the new iterate x_{k+1} cannot be uniquely determined either. As we show in proposition 3.3, descriptor systems or control systems may have the same behavior. One can interpret this as the input being implicit for descriptor systems. A descriptor system is defined as follows:

$$g(x_{k+1}) = f(x_k), \qquad x_k \in \mathcal{X} \tag{9}$$

where $f, g : \mathbb{R}^n \to \mathbb{R}^r$. A trajectory of this system is any sequence $(x_k)_k$ that satisfies eq. (9). The integer r is usually smaller than n which gives a choice for x_{k+1} and breaks the unicity of trajectories starting at a given state.

Definition 2.2 (Robust invariant set)**.** A set $\mathcal{S} \subseteq \mathcal{X}$ is *robustly invariant* for the system (7) (resp. *strongly invariant* for the system (9)) if for any state $x_0 \in \mathcal{S}$, all trajectories of the system with initial state x_0 remain in $\mathcal{S}$.

Proposition 2.2. A set $\mathcal{S} \subseteq \mathcal{X}$ is robustly invariant for system (7), as defined in definition 2.2, if and only if

$$f(\mathcal{S}, \mathcal{W}) \subseteq \mathcal{S}.$$

Proposition 2.3. A set $\mathcal{S} \subseteq \mathcal{X}$ is strongly invariant for system (9), as defined in definition 2.2, if and only if

$$g^{-1}(f(\mathcal{S})) \subseteq \mathcal{S}. \tag{10}$$

Definition 2.3 (Controlled invariant set)**.** A set $\mathcal{S} \subseteq \mathcal{X}$ is *controlled invariant* for system (8) (resp. *weakly invariant* for system (9)) if for any state $x_0 \in \mathcal{S}$, there exists a trajectory with initial state x_0 that remains in $\mathcal{S}$.

Proposition 2.4. A set $\mathcal{S} \subseteq \mathcal{X}$ is controlled invariant for system (8), as defined in definition 2.3, if and only if

$$\mathcal{S} \subseteq \pi_{[n_x],n_x+n_u}(f^{-1}(\mathcal{S}) \cap (\mathbb{R}^{n_x} \times \mathcal{U}))$$

where the π notation is defined in section 1.1

Proposition 2.5. A set $\mathcal{S} \subseteq \mathcal{X}$ is weakly invariant for system (9), as defined in definition 2.3, if and only if

$$f(\mathcal{S}) \subseteq g(\mathcal{S}).$$

Remark 2.1. The difference between system (7) and system (8) is the nature of the external input w or u. The disturbance w has an adversarial nature, it is considered arbitrarily chosen in $\mathcal{W}$ and cannot be controlled. On the other hand, we can choose the control u in $\mathcal{U}$ that fits our purpose. For this reason, the quantifier used in definition 2.2 is "all" while the quantifier used in definition 2.3 is "exists". A similar distinction applies for descriptor systems. For strong invariance, the choice is arbitrary while for weak invariance, we can choose the vector x_{k+1} among all of which have the same image through g.

We have shown how to formulate the invariance of sets as an inclusion. Any set satisfying the inclusion constraint is therefore guaranteed to be invariant although much larger invariant sets exists. Alternatively, one may be interested in a set that contains all invariant sets. This is out of the scope of this chapter, the reader is referred to [22] for an approach for such computation for control systems.

2.3.1 Switched Systems

In the previous section, the state space $\mathcal{X}$ considered is a subset of $\mathbb{R}^n$ which is usually convex. However, complex modern applications also require modeling a state q that belongs to a union of discrete values [20, 33]. We refer to x as the continuous part and q as the discrete part of the state space. A system with a purely continuous state space is adequately modeled by one of the systems defined in the previous section.

When the state space has both a continuous and discrete part, we say that it is *hybrid*. Arguably the simplest model of hybrid system is the discrete-time switched system [24]. A discrete-time switched system is characterized by a finite set of functions $\{f_1, f_2, \dots, f_m\}$ and the iteration

$$x_{k+1} = f_{\sigma_k}(x_k), \qquad \sigma_k \in [m] \tag{11}$$

where $[m]$ denotes the set $\{1, \dots, m\}$. It is important to distinguish three types of switching: *autonomous switching*, *controlled switching* and *stochastic switching*; see details on the difference between autonomous switching and controlled switching in [33, Section 1.1.3].

Proposition 2.1 is generalized as follows for switching systems with autonomous switching.

Proposition 2.6. A set $\mathcal{S}$ is invariant for system (11), as defined in definition 2.1, if and only if

$$f_\sigma(\mathcal{S}) \subseteq \mathcal{S}, \qquad \forall \sigma \in [m].$$

Similarly, we can define a switching version of each system of the previous section and generalize the formulation of invariance as m inclusion constraints. While it is relatively straightforward to generalize the set program for switched systems, the resulting set program may be considerably more difficult to solve. For instance, it is well known that it is not conservative to consider the ellipsoidal template for linear discrete-time autonomous system (6) and computing such ellipsoid can be achieved by solving a system of linear equations [37]. However, for linear discrete-time autonomous switched system (11), this template is conservative and the existence of a solution to the set program given in proposition 2.6 is even undecidable [13].

Remark 2.2. Since many control problems require the computation of nonconvex sets, parts of this chapter may seem restricted due to the convexity of the sets. Indeed, section 3.3 and section 5 are restricted to set programs for which the set decision variables $\mathcal{S}_i$ are convex bodies. Moreover, all templates presented in this paper except the polyset template (see section 4.4) are families of convex sets.

However, the existence of a convex invariant set is equivalent to stability for *linear* switched systems [24, Theorem 2.2] although stability can also be guaranteed by nonconvex sets as illustred by example 4.7. Besides, the maximal invariant set of linear systems (possibly controlled or switched) is the convex hull of the union of all invariant sets and is hence convex.

In the case of *nonlinear* systems, the convexity also appears necessary for proving the stability of systems switching in a convex set of dynamics (as opposed to a finite set of m dynamics as in (11)) [1, Section IV]. Finally, set programs with convex set decision variables can be used to compute nonconvex sets via *path-complete methods* [2]. These nonconvex sets are obtained by combining the convex sets with union and intersection operations which are encoded in the combinatorial structure of the inclusion constraints of the set program via the *path-complete framework* [41].

3 Template-independent set programming

3.1 Invariance inclusions

In this section, we focus on inclusion constraints that are in either of the two following forms:

$$f(\mathcal{S}_1) \subseteq \mathcal{S}_2, \tag{12}$$

$$\mathcal{S}_1 \subseteq f^{-1}(\mathcal{S}_2). \tag{13}$$

which are obtained by reformulating invariance constraints of autonomous systems as shown in proposition 2.1 and proposition 2.6. Note that these two inclusions are equivalent by proposition A.1.

We show in theorem 3.1 that for a general nonlinear function f, we can formulate the inclusions of sublevel sets of given functions as an inequality in term of these functions. If the function f is linear, we show in theorem 3.2 that we can formulate inequalities in terms of either the support (resp. gauge) function of $\mathcal{S}_1$ and $\mathcal{S}_2$ with (15) (resp. (16)) or we can write them in terms of the gauge (resp. support) function of the polar of $\mathcal{S}_1$ and $\mathcal{S}_2$ with (17) (resp. (18)).

Theorem 3.1. Consider a positive real number $\alpha > 0$, positively homogeneous functions $g_i : \mathbb{R}^{n_i} \to \mathbb{R}$ the sets $\mathcal{S}_i = \{\, x \in \mathbb{R}^{n_i} \mid g_i(x) \leq \alpha \,\}$ for $i = 1, 2$ and a function $f : \mathbb{R}^{n_1} \to \mathbb{R}^{n_2}$. The constraint (12) (resp. (13)) is equivalent to the following inequality:

$$\forall x \in \mathbb{R}^{n_1}, \qquad g_1(x) \geq g_2(f(x)). \tag{14}$$

Proof. Using (13), proposition A.2 and proposition A.3, we obtain (14). □

Theorem 3.2. Consider closed convex sets $\mathcal{S}_i \subseteq \mathbb{R}^{n_i}$ containing the origin for $i = 1, 2$ and a function $f(x) = Ax$ for some matrix $A \in \mathbb{R}^{n_2 \times n_1}$. The constraint (12) (resp. (13)) is equivalent to each of the following equivalent inequalities:

$$\forall y \in \mathbb{R}^{n_2}, \qquad \delta^*(A^\top y|\mathcal{S}_1) \leq \delta^*(y|\mathcal{S}_2). \tag{15}$$

$$\forall x \in \mathbb{R}^{n_1}, \qquad \mathrm{g}(\mathcal{S}_1, x) \geq \mathrm{g}(\mathcal{S}_2, Ax). \tag{16}$$

$$\forall y \in \mathbb{R}^{n_2}, \qquad \mathrm{g}(\mathcal{S}_1^\circ, A^\top y) \leq \mathrm{g}(\mathcal{S}_2^\circ, y). \tag{17}$$

$$\forall x \in \mathbb{R}^{n_1}, \qquad \delta^*(x|\mathcal{S}_1^\circ) \geq \delta^*(Ax|\mathcal{S}_2^\circ). \tag{18}$$

where g denotes the gauge function (definition A.2) and $\delta^*(\cdot|\cdot)$ denotes the support function (definition A.5).

Proof. Using (12) and proposition A.25, we obtain (15). Using (13) and proposition A.24, we obtain (16). Using (15) and proposition A.4, we obtain (17). Using (16) and proposition A.4, we obtain (18). □

Continuous-time results corresponding to theorem 3.2 are developed in [21].

3.2 Strong invariance inclusions

In this section, we show how substitutions of set variables into preimages of new set variables allow to rewrite the constraints into equivalent forms using the properties of set-valued maps reviewed in appendix A.1. We show an application of such substitutions for the computation of strongly invariant sets of a descriptor system.

The following proposition shows how to simplify an inner image with a substitution.

Proposition 3.1. Consider sets $\mathcal{S}_1 \subseteq \mathbb{R}^{n_1}, \mathcal{S}_2 \subseteq \mathbb{R}^{n_2}, \mathcal{S}_3 \subseteq \mathbb{R}^{n_3}$ and functions $f : \mathbb{R}^{n_1} \to \mathbb{R}^{n_3}, g : \mathbb{R}^{n_3} \to \mathbb{R}^{n_2}$. If $\mathcal{S}_1 = f^{-1}(\mathcal{S}_3)$ then the inclusion

$$g(f(\mathcal{S}_1)) \subseteq \mathcal{S}_2$$

is equivalent to the inclusion

$$g(\mathcal{S}_3) \subseteq \mathcal{S}_2.$$

Proof. Follows from eq. (50). □

The following proposition shows how to simplify an outer preimage with a substitution.

Proposition 3.2. Consider sets $\mathcal{S}_1 \subseteq \mathbb{R}^{n_1}, \mathcal{S}_2 \subseteq \mathbb{R}^{n_2}, \mathcal{S}_3 \subseteq \mathbb{R}^{n_3}$ and functions $f : \mathbb{R}^{n_2} \to \mathbb{R}^{n_3}, g : \mathbb{R}^{n_1} \to \mathbb{R}^{n_3}$. If $\mathcal{S}_2 = f^{-1}(\mathcal{S}_3)$ then the inclusion

$$f^{-1}(g(\mathcal{S}_1)) \subseteq \mathcal{S}_2$$

is equivalent to the inclusion

$$g(\mathcal{S}_1) \subseteq \mathcal{S}_3.$$

Proof. Follows from eq. (53). □

The following theorem shows how to combine these substitutions in order to compute strongly invariant sets for descriptor systems. This generalizes [38, Corollary 3.2], which studies the particular case of linear systems with the ellipsoidal template. The corresponding result for continuous-time systems is studied in [34, Theorem 2.7].

Theorem 3.3. Consider a set $\mathcal{T} \subseteq \mathbb{R}^r$. The sets $f^{-1}(\mathcal{T})$ and $g^{-1}(\mathcal{T})$ are strongly invariant for system (9), as defined in definition 2.2, if and only if

$$g^{-1}(\mathcal{T}) \subseteq f^{-1}(\mathcal{T}). \tag{19}$$

Proof. By proposition 3.1 (resp. proposition 3.2), if $\mathcal{S} = f^{-1}(\mathcal{T})$ (resp. $\mathcal{S} = g^{-1}(\mathcal{T})$) then eq. (10) is equivalent to eq. (19). □

3.3 Controlled invariance inclusions for convex bodies

In this section, we study inclusion constraints that are in either of the two following forms:

$$\mathcal{S}_1 \subseteq \pi_{[n_1],n_1+m} \begin{bmatrix} A & B \end{bmatrix}^{-1} \mathcal{S}_2, \tag{20}$$

$$C\mathcal{S}_1 \subseteq E\mathcal{S}_2. \tag{21}$$

with $\mathcal{S}_1 \subseteq \mathbb{R}^{n_1}, \mathcal{S}_2 \subseteq \mathbb{R}^{n_2}, A \in \mathbb{R}^{n_2 \times n_1}, B \in \mathbb{R}^{n_2 \times m}$, $C \in \mathbb{R}^{r \times n_1}$, $E \in \mathbb{R}^{r \times n_2}$ and the π notation is defined in section 1.1. By proposition 2.4, (20) is obtained by formulating controlled invariance for control systems (8) with no input constraints. By proposition 2.5, (21) is obtained by formulating controlled invariance for descriptor systems (9). This is illustrated in example 3.1,

Remark 3.1. It is important to distinguish this section with [26]. On the one hand, the approach presented in this section computes *convex* sets that are guaranteed to be controlled invariant. It applies to systems with a dynamics that must be linear although it can be hybrid [29]. On the other hand, the approach of [26] provides a hierarchy of possibly *nonconvex* sets that are not controlled invariant but converge to the maximal controlled invariant set. It applies to systems with a polynomial dynamic.

The following proposition shows the equivalence between both inclusion constraints, allowing us to focus on (21) without loss of generality.

Lemma 3.1. Given a subset $\mathcal{S} \subseteq \mathbb{R}^n$ and matrices $A \in \mathbb{R}^{r \times n}, B \in \mathbb{R}^{r \times m}$, the following holds:

$$A\mathcal{S} + B\mathbb{R}^m = \pi_{\mathrm{Im}(B)^\perp}^{-1} \pi_{\mathrm{Im}(B)^\perp} A\mathcal{S}$$

where $\pi_{\mathrm{Im}(B)^\perp}$ is an orthogonal projection matrix with range $\mathrm{Im}(B)^\perp$, the orthogonal complement of the linear subspace $\mathrm{Im}(B)$.

Proof. Given $x \in \mathcal{S}$ and $y \in \mathbb{R}^r$, we have $y \in A\{x\} + B\mathbb{R}^m$ if and only if $y - Ax \in \mathrm{Im}(B)$. As $\pi_{\mathrm{Im}(B)^\perp}$ is orthogonal, its kernel is $\mathrm{Im}(B)$. Therefore $y - Ax \in \mathrm{Im}(B)$ is equivalent to $\pi_{\mathrm{Im}(B)^\perp} y = \pi_{\mathrm{Im}(B)^\perp} Ax$. □

Proposition 3.3. Consider sets $\mathcal{S}_i \subseteq \mathbb{R}^{n_i}$ for $i = 1, 2$. The inclusion (20) is equivalent to the inclusion (21) where $C = \pi_{\mathrm{Im}(B)^\perp} A$ and $E = \pi_{\mathrm{Im}(B)^\perp}$.

Proof. A vector $x \in \pi_{[n_1],n_1+m} \begin{bmatrix} A & B \end{bmatrix}^{-1} \mathcal{S}_2$ if and only if there exists a vector u such that $Ax + Bu \in \mathcal{S}_2$. This is equivalent to $(A\{x\} + B\mathbb{R}^m) \cap \mathcal{S}_2 \neq \emptyset$. By lemma 3.1, this can be rewritten as

$$\pi_{\mathrm{Im}(B)^\perp}^{-1} \pi_{\mathrm{Im}(B)^\perp} Ax \cap \mathcal{S}_2 \neq \emptyset$$

or equivalently

$$\pi_{\mathrm{Im}(B)^\perp} Ax \cap \pi_{\mathrm{Im}(B)^\perp} \mathcal{S}_2 \neq \emptyset.$$

Therefore, (20) is equivalent to

$$\pi_{\mathrm{Im}(B)^\perp} A\mathcal{S}_i \subseteq \pi_{\mathrm{Im}(B)^\perp} \mathcal{S}_2.$$

□

Example 3.1. Consider the discrete-time linear control system, as defined in see (8), $x_{k+1} = x_k + u_k/2$ where the state x_k and the control u_k are constrained to belong to the closed interval $\mathcal{X} = \mathcal{U} = [-1, 1]$. Note that a set is controlled invariant for this system if and only if it is the projection onto the first dimension of an controlled invariant set for the following system

$$\begin{aligned} x_{k+1} &= x_k + \frac{u_k}{2} \\ u_{k+1} &= u'_k \end{aligned}$$

where the state vector (x_k, u_k) is constrained to belong to the square $\mathcal{X} = [-1, 1]^2$ and the control u'_k is unconstrained, i.e. $\mathcal{U} = \mathbb{R}$. By proposition 2.4, a set $\mathcal{S}$ is controlled invariant for this system if and only if

$$\mathcal{S} \subseteq \pi_{[2],3} A^{-1} \mathcal{S} \tag{22}$$

where

$$A = \begin{bmatrix} 1 & 1/2 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Now consider the discrete-time linear descriptor system, as defined in (9), with states (x_k, u_k) and difference equation $x_{k+1} = x_k + \frac{u_k}{2}$. Note that u_{k+1} can be chosen independently of the value of x_k and u_k. By proposition 2.5, a set $\mathcal{S}$ is weakly invariant for this system if and only if

$$\begin{bmatrix} 1 & 1/2 \end{bmatrix} \mathcal{S} \subseteq \begin{bmatrix} 1 & 0 \end{bmatrix} \mathcal{S}. \tag{23}$$

By proposition 3.3, (23) and (22) are equivalent, which illustrates the equivalence between the controlled invariance of control systems and the weak invariance of descriptor systems.

To show the relevance of proposition 3.3, we argue below that the inclusion (20) does not seem to be easily rewritten in terms of the gauge or support function of $\mathcal{S}_1$ and $\mathcal{S}_2$. First note that as there is both an image and a preimage in the right-hand side. Therefore, we cannot get an inequality in terms of $\mathrm{g}(\mathcal{S}_2, \cdot)$ because of the image and cannot get an inequality in terms of $\delta^*(\cdot|\mathcal{S}_2)$ because of the preimage. On the other hand, the inclusion $\pi_{[n_1],n_1+m}^{-1} \mathcal{S}_1 \subseteq \begin{pmatrix} A & B \end{pmatrix}^{-1} \mathcal{S}_2$ can be rewritten as an inequality in terms of $\mathrm{g}(\mathcal{S}_1, \cdot)$ and $\mathrm{g}(\mathcal{S}_2, \cdot)$ using proposition A.24 but this inclusion is not equivalent to the inclusion (20) as we saw in Remark A.1.

By proposition A.26, the inclusion (20) is equivalent to

$$\mathcal{S}_1^\circ \supseteq \pi_{[n_1],n_1+m}^{-1} \begin{pmatrix} A & B \end{pmatrix}^\top \mathcal{S}_2^\circ \tag{24}$$

This inclusion is generalized in [44] for robustly controlled invariant sets of discrete-time control linear systems with disturbances.

Again, there is both an image and a preimage in the right-hand side of the inclusion (24), and it is not equivalent to $\pi_{[n_1],n_1+m}^\top \mathcal{S}_1^\circ \supseteq \begin{pmatrix} A & B \end{pmatrix}^\top \mathcal{S}_2^\circ$ by Remark A.1. Hence there does not seem to be any way to use this inclusion to write an inequality in terms of either the gauge or support function of $\mathcal{S}_1^\circ$ and $\mathcal{S}_2^\circ$.

We now consider the inclusion (21). This inclusion can be rewritten in terms of the support function of $\mathcal{S}_1$ and $\mathcal{S}_2$ or the Minkowski function of $\mathcal{S}_1^\circ$ and $\mathcal{S}_2^\circ$ as shown in the following theorem.

Theorem 3.4. Consider nonempty closed convex sets $\mathcal{S}_1$, $\mathcal{S}_2$, the inclusion (21) is equivalent to the following equivalent inequalities

$$\forall y \in \mathbb{R}^r, \quad \delta^*(C^\top y|\mathcal{S}_1) \leq \delta^*(E^\top y|\mathcal{S}_2) \tag{25}$$

$$\forall y \in \mathbb{R}^r, \quad \mathrm{g}(\mathcal{S}_1^\circ, C^\top y) \leq \mathrm{g}(\mathcal{S}_2^\circ, E^\top y). \tag{26}$$

where g denotes the gauge function (definition A.2) and $\delta^*(\cdot|\cdot)$ denotes the support function (definition A.5).

Proof. By proposition A.10, (21) is equivalent to

$$\forall y \in \mathbb{R}^r, \quad \delta^*(y|C\mathcal{S}_1) \leq \delta^*(y|E\mathcal{S}_2).$$

By proposition A.25, this inequality is equivalent to (25). By proposition A.4, the inequalities (25) and (26) are equivalent. □

3.4 Fixed point reformulation of set programs

In this section, we investigate the relation between some monotonicity property (defined in appendix A.1) of a set program and the ability to solve it by iterative set transformations until convergence. Similar set propagation techniques are also used in reachability analysis as reviewed in [3].

Consider a set program of the form:

$$\begin{aligned} \underset{\mathcal{S}_1 \subseteq \mathbb{R}^{n_1}, \ldots, \mathcal{S}_N \subseteq \mathbb{R}^{n_N}}{\text{maximize}} \quad & f(\mathcal{S}_1, \ldots, \mathcal{S}_N) \\ \text{subject to:} \quad & \mathcal{S}_i \subseteq h_i(\mathcal{S}_1, \ldots, \mathcal{S}_N), \quad i = 1, 2, \ldots, N \\ & \mathcal{S}_i \subseteq \mathcal{C}_i, \quad i = 1, 2, \ldots, N. \end{aligned} \tag{27}$$

Theorem 3.5. Consider a set program of the form (27) and the iteration

$$\mathcal{S}_i^{k+1} = \mathcal{S}_i^k \cap h_i(\mathcal{S}_1^k, \ldots, \mathcal{S}_N^k) \tag{28}$$

with $\mathcal{S}_i^0 = \mathcal{C}_i$ for all $i \in [N]$. If f is monotone and the functions h_i are monotone for all $i \in [N]$ then the iteration (28) converges to an optimal solution to the set program.

Proof. By (28), the sequence is monotone hence it converges to sets $\mathcal{S}^\infty$. As $\mathcal{S}^\infty$ satisfies eq. (28) by equality, it is feasible for the set program.

Let $\mathcal{S}_i^\star$ be any optimal solution to the set program. To show that the objective function f is not smaller for $\mathcal{S}^\infty$ than for $\mathcal{S}^\star$, it is sufficient to prove that $\mathcal{S}_i^\infty \supseteq \mathcal{S}_i^\star$ for all $i \in [N]$ by the monotonicity of f. To this end, we prove by induction that $\mathcal{S}_i^k \supseteq \mathcal{S}_i^\star$ for all $i \in [N]$. The case $k = 0$ follows from the feasibility of $\mathcal{S}^\star$ for the set program. Suppose it is the case for a given k. By the monotonicity of h_i, we have

$$h_i(\mathcal{S}_1^k, \ldots, \mathcal{S}_N^k) \supseteq h_i(\mathcal{S}_1^\star, \ldots, \mathcal{S}_N^\star) \supseteq \mathcal{S}_i^\star.$$

This shows that $\mathcal{S}_i^{k+1} \supseteq \mathcal{S}_i^\star$. □

Remark 3.2. Note that the the iteration (28) does not depend on the objective function f. For this reason, this optimal solution is commonly referred to as the *maximal* solution without specifying any objective, as it is optimal for any monotone objective.

We have a similar result for set programs of the following form by replacing the intersection by a union in the iteration.

$$\begin{aligned} \underset{\mathcal{S}_1\subseteq\mathbb{R}^{n_1},\ldots,\mathcal{S}_N\subseteq\mathbb{R}^{n_N}}{\text{minimize}} \quad & f(\mathcal{S}_1,\ldots,\mathcal{S}_N) \\ \text{subject to:} \quad & h_i(\mathcal{S}_1,\ldots,\mathcal{S}_N)\subseteq\mathcal{S}_i, \quad i=1,2,\ldots,N \\ & \mathcal{C}_i\subseteq\mathcal{S}_i, \quad i=1,2,\ldots,N. \end{aligned} \tag{29}$$

Theorem 3.6. Consider a set program of the form (29) and the iteration

$$\mathcal{S}_i^{k+1}=\mathcal{S}_i^k\cup h_i(\mathcal{S}_1^k,\ldots,\mathcal{S}_N^k) \tag{30}$$

with $\mathcal{S}_i^0=\mathcal{C}_i$ for all $i\in[N]$. If f is monotone and the functions h_i are monotone for all $i\in[N]$ and the set program has a bounded optimal solution $\mathcal{S}^\star$ then the iteration (28) converges to an optimal solution to the set program.

4 Four numerical templates

In this section, we discuss the numerical computation of solutions to set programs with an emphasis on the impact of the choice of representation to the tractability of the computation. We illustrate that the choice between the representation with a gauge or support function is essentially template-independent with the following example.

Example 4.1. Consider the following generic set program modeling controlled invariance considered in example 3.1:

$$\underset{\mathcal{S}}{\text{maximize}}\, f(\mathcal{S}) \tag{31}$$

$$C\mathcal{S}\subseteq E\mathcal{S} \tag{32}$$

$$\mathcal{S}\subseteq[-1,1]^2. \tag{33}$$

where $C=\begin{bmatrix}1 & 1/2\end{bmatrix}$ and $E=\begin{bmatrix}1 & 0\end{bmatrix}$ (see eq. (23)) and f is any monotone function. As discussed in remark 3.2, since the set program satisfies the requirements of theorem 3.5, the limit of the iteration (28) is optimal for any monotone function f.

Template	$\cap$	conv $\cup$	$+$	$\sharp$	A	A^{-1}	$\circ$
Polyhedra	✓	✓	✓	✓	✓	✓	✓
Zonotopes	✗	✗	✓	✗	✓	✗	✗
Ellipsoids	✗	✗	✗	✗	✓	✓	✓
Polysets	✗	✗	✗	✗	✗	✓	✗
Piecewise semi-ellipsoids	✗	✗	✗	✗	✓	✓	✓
Piecewise polysets	✗	✗	✗	✗	✗	✓	✗

Table 1: This table highlights the invariance of the different templates studied in this chapter under the operations preserving convexity that we consider. Each row represents a template and each column represents an operation. The first symbol of the column denotes the operation being considered. Then, a ✓ symbol indicates that the template is invariant under the operation, a ✗ symbol indicates that it is not. A similar table can be found in [25, Table 1] or [3, Table 1].

4.1 Polyhedra

Polyhedra are ubiquitous in control. This can be in part attributed to the fact that this family is invariant under most operations of interest as we will see in this section; see also table 1. However, polyhedra possess two dual representations and each operation is usually only computationally cheap in one of them. Also, converting to a dual representation is computationally expensive in high dimension as discussed in appendix A.4.3. We will see that this is closely related to table 2, as suggested by proposition A.27 and proposition A.30.

4.1.1 Polyhedral computation of set programs: fixed point approach

The approach described in section 3.4 is well suited for polyhedra as this family is invariant under various operations. Moreover, depending on the operations, the choice between $\mathcal{H}$-representation, $\mathcal{V}$-representation or $\mathcal{Z}$-representation is crucial to avoid having to compute a representation conversion or polyhedral projection of $\mathcal{H}$-representation.

For instance, for set programs of the form (27), as there is an intersection in the iteration (28), the $\mathcal{H}$-representation is used in practice in view of proposition A.28.

On the other hand, for set programs of the form (29), there is a union in the iteration (30). As the template of polyhedra only contains convex sets, the union can be replaced by the convex hull of the union. In view of proposition A.31, the $\mathcal{V}$-representation is used in practice for such programs.

As proposition A.28 and proposition A.31 may generate redundant elements in the representation, it is crucial to remove redundant ones after each iteration. We can determine whether each element is redundant by solving a linear program for each element as shown in proposition A.35, proposition A.37 and proposition A.39.

For some set programs, we may still resort to using a representation conversion or polyhedral projection in each iteration. For instance, for the computation of a controlled invariant set of a discrete-time linear control system, the inclusion given in proposition 2.4 involves a projection, linear preimage and an intersection.

Example 4.2. Consider the set program of example 4.1 with the constraint eq. (32) written in the form eq. (22) to satisfy the form of set program (27). Following theorem 3.5, the optimal solution can be obtained by the iteration

$$\mathcal{S}^{k+1} = \mathcal{S}^k \cap \pi_{[2],3} A^{-1} \mathcal{S}$$

with $\mathcal{S}^0 = [-1, 1]^2$. Given a $\mathcal{H}$-representation of $\mathcal{S}$, a $\mathcal{H}$-representation of $A^{-1}\mathcal{S}$ can be obtained with proposition A.29. Then $\pi_{[2],3} A^{-1}\mathcal{S}$ can be obtained either through projection, e.g., with Fourier-Motzkin elimination or block elimination. Then the intersection with $\mathcal{S}^k$ can be obtained with proposition A.28.

Alternatively, the $\mathcal{V}$-representation of $A^{-1}\mathcal{S}$ can be computed, and then the projection of the $\mathcal{V}$-representation is obtained with corollary A.3 as discussed in appendix A.4.3. Then another representation conversion is carried out to obtain the $\mathcal{H}$-representation of $\pi_{[2],3} A^{-1}\mathcal{S}$. This requires one representation conversion in three dimensions and one in two dimensions. A computationally cheaper alternative is to compute the $\mathcal{H}$-representation of the $\pi_{[2],3} A^{-1}\mathcal{S}$ from the $\mathcal{H}$-representation $A^{-1}\mathcal{S}$ direction using block-elimination which only requires a representation conversion in one dimension.

It is recommended to remove redundancy in the $\mathcal{H}$-representation after each iteration. It is first used a convergence check and second, the time taken removing the redundancy is often largely outweighted by the time saved in the projection due to the reduced size of the representation.

This iteration actually converges in one iteration as we have $\mathcal{S}^2 = \mathcal{S}^1$. The optimal solution $\mathcal{S}^1$ is represented in yellow in fig. 4, fig. 6 and fig. 8.

In practice, $\mathcal{H}$-representations are used more frequently than $\mathcal{V}$-representations. This may seem surprising as the $\mathcal{V}$-representation of the polar of a set $\mathcal{H}$-representation of the primal set, as shown by proposition A.33. Hence for every application for which it is more efficient to use $\mathcal{H}$-representations, there should be a corresponding "polar application" for which it is more efficient to use $\mathcal{V}$-representations.

Part of this gap may be attributed to the more frequent occurence of hypercubes compared to cross-polytopes (its polar by corollary A.4) in applications. Indeed, the $\mathcal{V}$-representation of hypercubes has a size exponential in the dimension while the $\mathcal{H}$-representation (or zonotope representation) of hypercubes has a size linear in the dimension.

It is important however to keep in mind that there are no theoretical reason why using the $\mathcal{H}$-representation would be more efficient in general. Therefore, given a set program, it is important to always consider both options and choose the most appropriate one.

As detailed in proposition A.31, the size of the $\mathcal{V}$-representation of a Minkowski sum grows quadratically, even if some of the vertices or rays in the resulting $\mathcal{V}$-representation may be redundant. On the other hand, as we showed in proposition A.41, the size of the $\mathcal{Z}$-representation of a Minkowski sum grows linearly.

For this reason, zonotopes is preferred over the $\mathcal{V}$-representation for solving set programs where $\mathcal{C}_i$ is a zonotope and h_i involves only linear image and Minkowski sums. This is the case for instance for the computation of controlled invariant sets for discrete-time linear system with disturbance defined by eq. (7) with $f(x_k, w_k) = Ax_k + Bw_k$ for some matrices A and B. If $\mathcal{W}$ is a zonotope, the inclusion constraint given in proposition 2.2 is well fitted for the $\mathcal{Z}$-representation.

4.1.2 Polyhedral computation of set programs: linear programming approach

The previous section shows how to compute the optimal solution of a set program of specific forms. This has however two drawbacks: First, each iteration may require polyhedral projection or representation conversion which is computationally costly. Second, the optimal solution may not be polyhedral or may be a polyhedron that has a $\mathcal{V}$-representation or $\mathcal{H}$-representation of large size.

It may therefore be desirable to search instead for a solution in a family of polyhedra with a representation of a given size. For instance, we may parametrize the search by the coordinates of s points (resp. s halfspaces). However, reformulating the inclusion constraints with proposition A.35 (resp. proposition A.39) gives rise to inequalities that are bilinear in the coordinates of the points (resp. halfspaces) and of the multipliers λ.

One way to circumvent this is to fix the pieces of the piecewise linear gauge or support function of the polyhedron [43, 44].

Definition 4.1 ([31, Definition 2])**.** A *polyhedral conic partition* of $\mathbb{R}^n$ is a set of m polyhedral cones $(\mathcal{P}_i)_{i=1}^m$ with nonempty interior such that for all $i \neq j$, $\dim(\mathcal{P}_i \cap \mathcal{P}_j) < n$ and $\cup_{i=1}^m \mathcal{P}_i = \mathbb{R}^n$.

A polyhedral conic partition defines the full-dimensional faces of a *complete fan*, as defined in [50, Section 7]. Given a polyhedral conic partition, we use the notation

$$\mathcal{N} = \{\, (i,j) \mid \dim(\mathcal{P}_i \cap \mathcal{P}_j) = n-1 \,\}.$$

For each $(i,j) \in \mathcal{N}$, the affine hull, of $\mathcal{P}_i \cap \mathcal{P}_j$ is a hyperplane. We denote by n_{ij} the normal of this hyperplane directed towards $\mathcal{P}_i$, i.e., such that $\mathcal{P}_i \subseteq \mathcal{H}_{-n_{ij},0}$.

Given the $\mathcal{H}$-representation of a polytope $\mathcal{S}$, let $\mathcal{F}(\mathcal{S})$ be the conic partition obtained by considering the conic hull of each facets of $\mathcal{S}$; see [43, Proposition 5]. This is commonly referred to as the *face fan* of $\mathcal{S}$ [50, Example 7.2]. For instance $\mathcal{F}([-1,1]^2)$ gives the four quadrants as cones of the conic partition.

Given a polyhedral conic partition, the $\mathcal{H}$-representation of a polytope $\mathcal{S}$ can then be parametrized by vectors $(a_i, 1)_{i=1}^m \in \mathcal{H}_{\mathrm{rep}}(\mathcal{S})$ such that (see proposition A.27):

$$\mathrm{g}(\mathcal{S}, x) = \max_{i \in [m]} \langle a_i, x \rangle = \langle a_i, x \rangle \text{ if } x \in \mathcal{P}_i. \tag{34}$$

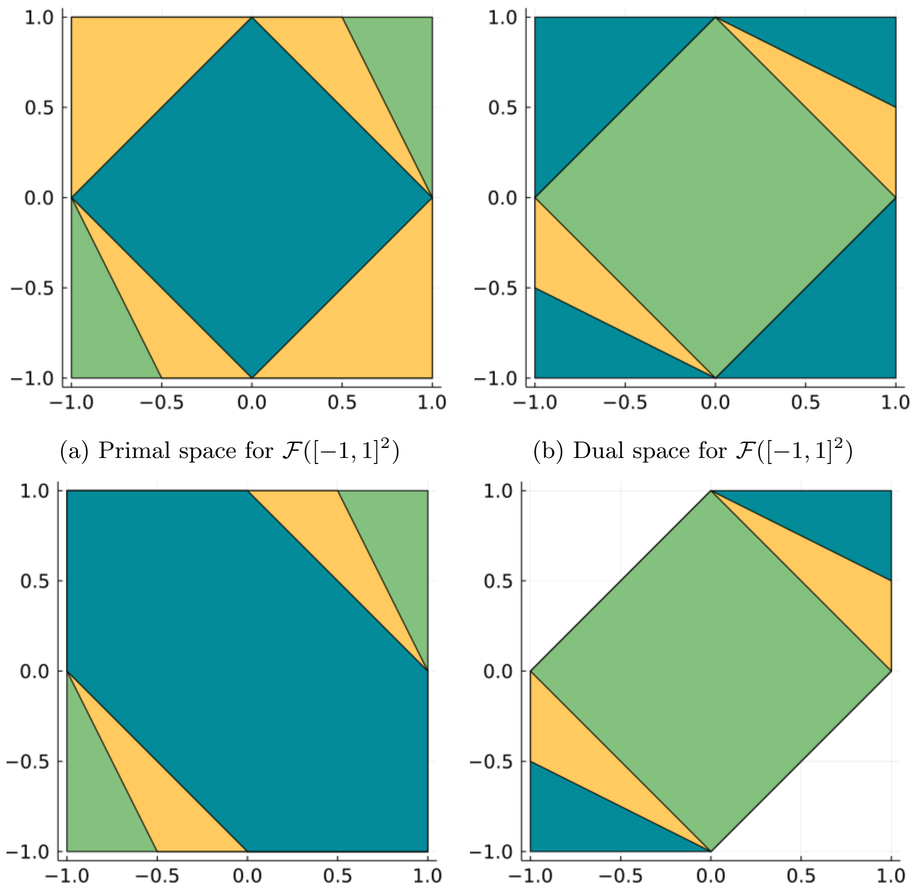
Similarly, the $\mathcal{V}$-representation of a polytope $\mathcal{S}$ can then be parametrized by vectors $((a_i)_{i=1}^m, ()) \in \mathcal{V}_{\mathrm{rep}}(\mathcal{S})$ (where $()$ denotes the absence of rays) such that (see proposition A.30):

$$\delta^*(x|\mathcal{S}) = \max_{i \in [m]} \langle a_i, x \rangle = \langle a_i, x \rangle \text{ if } x \in \mathcal{P}_i. \tag{35}$$

Both eq. (34) and eq. (35) are ensured by the constraints, $\forall (i,j) \in \mathcal{N}$,

$$\begin{aligned} \exists \lambda \in \mathbb{R}, \quad a_i - a_j &= \lambda n_{ij} \\ \langle a_i, n_{ij} \rangle &\geq \langle a_j, n_{ij} \rangle. \end{aligned}$$

Example 4.3 illustrates the computation of polytopes using this technique.



(a) Primal space for $\mathcal{F}([-1,1]^2)$ (b) Dual space for $\mathcal{F}([-1,1]^2)$

(c) Primal space for $\mathcal{F}(\mathrm{conv}([-1,1]^2, 1.25([-1,1]^2)^\circ))$ (d) Dual space for $\mathcal{F}(\mathrm{conv}([-1,1]^2, 1.25([-1,1]^2)^\circ))$

Figure 3: Optimal solutions for the set program of Example 4.1 for polytopes with a fixed face fan. The green set is the square $[-1,1]^2$ and the blue and yellow sets are the optimal solutions with the objective replaced by the integral of $\delta^*(y|\mathcal{S})^2$ over the polar of the square and with different face fans.

Example 4.3. Consider the set program defined in example 4.1. For the polyhedral template, there are several choices of parametrization of the representation. We consider two cases here and highlight that in both cases, it is more appropriate to use the dual representation, that is, the support function of $\mathcal{S}$ which is given by the $\mathcal{V}$-representation of $\mathcal{S}$ or equivalently the $\mathcal{H}$-representation of $\mathcal{S}^\circ$. This supports the template-independence of this choice of representation.

The $\mathcal{V}$-representation of both $C\mathcal{S}$ and $E\mathcal{S}$ depends linearly on the $\mathcal{V}$-representation of $\mathcal{S}$ as shown in proposition A.32. The inclusion is ensured by constraining each element of the $\mathcal{V}$-representation of $C\mathcal{S}$ to belong to $E\mathcal{S}$ using proposition A.35 and proposition A.37. If the coordinates of each element of the $\mathcal{V}$-representation of $\mathcal{S}$ is fixed, this gives us a linear program to verify that $\mathcal{S}$ is a feasible solution. Otherwise, if the coordinates of the elements are to be determined it results in a bilinear program. This condition is given in [10, Theorem 4.44]. Note that there is no corresponding result for the $\mathcal{H}$-representation in [10]. This is in accordance with the claim of Section 3.3 that the dual representation should be used for such inclusion.

Alternatively, given a fixed face fan, consider a parametrization of the support function as in eq. (35). In this case, by proposition A.25, we have

$$\delta^*(y|C\mathcal{S}) = \langle Ca_i, y\rangle \text{ if } y \in C^\top \mathcal{P}_i.$$

Therefore, $C\mathcal{S} \subseteq E\mathcal{S}$ is rewritten as

$$\forall i, j, \langle Ea_j, x\rangle \geq \langle Ca_i, x\rangle, \forall x \in C^\top \mathcal{P}_i \cap E^\top \mathcal{P}_j. \tag{36}$$

or equivalently

$$\forall i, j, Ea_j - Ca_i, (C^\top \mathcal{P}_i \cap E^\top \mathcal{P}_j)^*. \tag{37}$$

If the face fan is fixed, this is rewritten as linear constraints using either proposition A.39 with eq. (36) or proposition A.35 with eq. (37); both giving the same linear programming formulation.

For the fan $\mathcal{F}(\mathcal{S}^2)$ where $\mathcal{S}^2$ is the polytope found in example 4.2, the optimal solution is $\mathcal{S}^2$. For the fan $\mathcal{F}(([-1,1]^2)^\circ)$, the optimal solution is $[-1,1] \times \{0\}$. The solutions for other fans are illustrated in fig. 3.

4.2 Ellipsoids

An *ellipsoid* is defined by

$$\mathcal{E}_P = \{\, x \in \mathbb{R}^n \mid x^\top P x \leq 1 \,\}. \tag{38}$$

where P is a positive definite matrix. When P is only positive semidefinite, the set is a *semi-ellipsoid*.

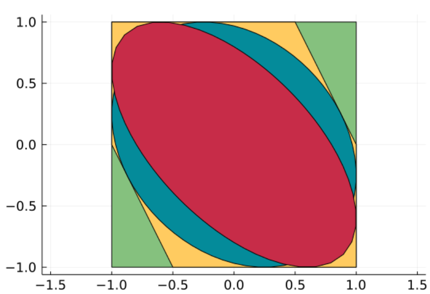
The gauge function and gauge-like function of degree 2 (see definition A.2) of the ellipsoid are

$$\mathrm{g}(\mathcal{E}_P, x) = \sqrt{x^\top P x} \qquad\qquad \mathrm{g}_2(\mathcal{E}_P, x) = x^\top P x/2. \tag{39}$$

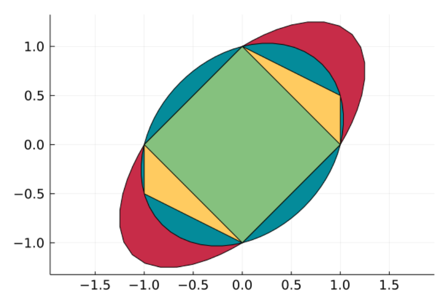
Example 4.4. For the ellipsoids template, the dual representation is given by $\delta^*(y|\mathcal{S}) = \sqrt{x^\top Q x}$ as shown in proposition A.45. By (57), the inclusion (33) is equivalent to $(1,1), (-1,1), (1,-1), (-1,-1) \in \mathcal{S}^\circ = \mathcal{E}_Q$ which can be encoded as a linear constraint with (71). As we will see in Section 3.3, (32) can be encoded as the LMI:

$$CQ_iC^\top \preceq EQ_jE^\top.$$

In this case, as C, E have only one row, this LMI has dimension 1×1 so it is the linear inequality $Q_{1,2} + Q_{2,2}/4 \leq 0$. The optimal ellipsoid $\mathcal{E}_{Q^{-1}}$ and its polar $\mathcal{E}_Q$ are given respectively in Figure 4a and Figure 4b.



(a) Primal space



(b) Dual space

Figure 4: Optimal solutions for the set program of Example 4.1. The green set is the square $[-1,1]^2$, the yellow set is the optimal solution to the generic set program, the blue ellipsoid is the optimal solution to the set program for the ellipsoidal template and the red ellipsoid is the optimal solution to the set program for the ellipsoidal template with the objective replaced by the sum of squared length of the semi-axes of the polar.

4.3 Piecewise semi-ellipsoids

The conservatism of ellipsoidal controlled invariant sets and the complexity of the representation of polyhedral controlled invariant sets has motivated the search for alternative templates of sets. Moreover, for some sets, the set has a polyhedral boundary in some directions and a smooth boundary in other directions. Using different representations for these different directions would allow us to significantly improve our approximation ability for such sets. The template we study in this section is the family of piecewise semi-ellipsoids.

Piecewise semi-ellipsoids may either be defined as the 1-sublevel sets of piecewise quadratic forms or as sets with a piecewise semi-ellipsoidal Minkowski function.

Definition 4.2 (Piecewise semi-ellipsoidal sets)**.** A closed convex subset $\mathcal{S} \subseteq \mathbb{R}^n$ containing the origin is said to be *piecewise semi-ellipsoidal* if there exists a polyhedral conic partition $(\mathcal{P}_i)_{i=1}^m$ and symmetric positive semidefinite matrices $Q_i \in \mathcal{S}_+^n$ for $i = 1, \ldots, m$, such that

$$
\begin{aligned}
&\mathrm{g}(\mathcal{S}, x) = \sqrt{x^\top Q_i x} && \text{if } x \in \mathcal{P}_i, \\
&x^\top Q_i x = x^\top Q_j x, \quad \forall i, j \in \mathcal{N}, x \in \mathcal{P}_i \cap \mathcal{P}_j && (40)\\
&n_{ij}^\top Q_i x \geq n_{ij}^\top Q_j x, \quad \forall i, j \in \mathcal{N}, x \in \mathcal{P}_i \cap \mathcal{P}_j. && (41)
\end{aligned}
$$

Note that without (40) and (41), the 1-sublevel set of the piecewise semi-ellipsoidal function may be non-convex. The condition (40) ensures the continuity of the function and (41) ensures its convexity.

This family generalizes both ellipsoids and polyhedra. Indeed, if $m = 1$ and $\mathcal{P}_1 = \mathbb{R}^n$, we recover the family of ellipsoids and if the matrices Q_i are rank-1, we recover the family of polyhedra studied in section 4.1.2.

Remark 4.1. The intersection of ellipsoids is not necessarily a piecewise semi-ellipsoids. Indeed, consider a nonempty intersection of different ellipsoids $\mathcal{E}_{Q_1}$ and $\mathcal{E}_{Q_2}$. The partition should be defined according to the quadratic form $p(x) = x^\top P x$ where $P = Q_1 - Q_2$. In order for the partition to be polyhedral, we need the algebraic variety defined by the set of zeros of $p(x)$ to be the union of two (possibly identical) hyperplanes $\overline{\mathcal{H}}_{a_1,0}, \overline{\mathcal{H}}_{a_2,0}$. That is, we need $p(x) = \langle a_1, x\rangle\langle a_2, x\rangle$ or in other words, we need $P = (a_1 a_2^\top + a_2 a_1^\top)/2$. This holds if and only if P has rank 1 or 2. This shows that while the intersection of ellipsoids is always a piecewise semi-ellipsoid in the planar case, this does not necessarily hold in higher dimension.

Given a piecewise semi-ellipsoidal sets, its polar is also piecewise semi-ellipsoidal but the conic partition of the polar depends on the matrices Q_i. This is a consequence of proposition A.6 and the fact that the conjugate of a piecewise quadratic function is a piecewise quadratic function; see [45, Theorem 11.14]. The Minkowski function of the polar set has the closed form expression given by proposition 4.1.

Proposition 4.1. Given a piecewise semi-ellipsoidal set $\mathcal{S}$, as defined in definition 4.2, the polar set $\mathcal{S}^\circ$ is the piecewise semi-ellipsoidal representation with the polyhedral conic partition made of the polyhedral cones

$$\mathrm{conv}_{i \in I} Q_i \bigcap_{i \in I} \mathcal{P}_i \tag{42}$$

with matrices

$$E_I^\top (E_I Q_i E_I^\top)^\dagger E_I$$

where[1] $i \in I$ and $E_I = \pi_{\mathrm{aff} \cap_{i\in I} \mathcal{P}_i}$ for any subset I of $[m]$ such that the polyhedral cone given in eq. (42) has nonempty interior. In particular, for singletons $I = \{i\}$, E_I is the identity so this gives a polyhedral cone $Q_i \mathcal{P}_i$ with matrix $Q_i^\dagger$.

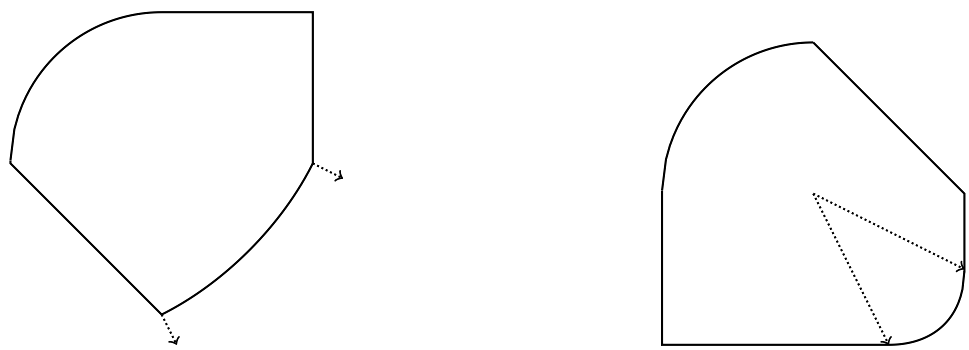
Proof. By proposition A.5, the polar of the gauge function of $\mathcal{S}$ is the support function of $\mathcal{S}$, that is, the gauge function of $\mathcal{S}^\circ$. Therefore, by proposition A.6, $\mathrm{g}_2(\mathcal{S}^\circ, x) = \mathrm{g}_2^*(\mathcal{S}, x)$. By corollary A.1, $y = \nabla \mathrm{g}_2(\mathcal{S}, x)$

[1] We have $E_I Q_i E_I^\top = E_I Q_j E_I^\top$ for any $i, j \in I$ by (40) so the matrix is independent on the $i \in I$ chosen.

if and only if $x = \nabla g_2^*(\mathcal{S}, x)$. Given $x \in \cap_{i \in I} \mathcal{P}_i$ where $I = \{ i \in [m] \mid x \in \mathcal{P}_i \}$, $\nabla g_2(\mathcal{S}, x) = \mathrm{conv}(\{ Q_i x \mid i \in I \})$. Since $x \in \cap_{i \in I} \mathcal{P}_i$, there exists z such that $x \in E_I^\top z$. Consider $y = Q_j x$ for some $j \in I$, we have $E_I y = E_I Q_j E_I^\top z$ hence $z = (E_I Q_j E_I^\top)^\dagger E_I y$. This implies that $x = E_I^\top z = E_I^\top (E_I Q_j E_I^\top)^\dagger E_I y$. By (40), $E_I Q_i E_I^\top = E_I Q_j E_I^\top$ for any $i, j \in I$ hence this holds for any $y \in \mathrm{conv}(\{ Q_i x \mid i \in I \})$. $\square$

Remark 4.2. The conic partition for the polar set created by proposition 4.1 seems to contain many polyhedral cones. This seems surprising as the polar operation is an involution for closed convex sets containing the origin. In fact, many polyhedral cones of the conic partition created can be dropped without changing the set as they are not full-dimensional. For instance, the pieces of the partition created in (42) are only full-dimensional in case the subdifferential of $g(\mathcal{S}, x)$ is not a singleton for all $x \in \cap_{i \in I} \mathcal{P}_i$. That is if the inequality (41) is not satisfied with equality for each pair of $i, j \in I$.

Example 4.5 illustrates the computation of the polar of a piecewise semi-ellipsoidal set with proposition 4.1.



(a) Set $\mathcal{S}$ defined by (43).

(b) Set $\mathcal{S}^\circ$ defined by (44).

Figure 5: Illustration for sets $\mathcal{S}$ and $\mathcal{S}^\circ$ defined in Example 4.5.

Example 4.5. Consider the piecewise semi-ellipsoidal set defined by

$$g(\mathcal{S}, x) = \begin{cases} |x| & \text{if } 0 \le y \le x, \\ |y| & \text{if } 0 \le x \le y, \\ \sqrt{x^2 + y^2} & \text{if } x \le 0 \le y, \\ |x + y| & \text{if } x, y \le 0, \\ \sqrt{x^\top Q_5 x} & \text{if } y \le 0 \le x. \end{cases} \tag{43}$$

where $x^\top Q_5 x = x^2 - xy + y^2$. The polar set is also piecewise semi-ellipsoidal:

$$g(\mathcal{S}^\circ, x) = \begin{cases} |x + y| & \text{if } 0 \le x, y, \\ \sqrt{x^2 + y^2} & \text{if } x \le 0 \le y, \\ |x| & \text{if } x \le y \le 0, \\ |y| & \text{if } y \le x, 2x + y \le 0, \\ \sqrt{x^\top Q_5^{-1} x} & \text{if } 2x + y \ge 0, x + 2y \le 0, \\ |x| & \text{if } x + 2y \ge 0, y \ge 0. \end{cases} \tag{44}$$

where $x^\top Q_5^{-1} x = (4/3) \cdot (x^2 + xy + y^2)$.

Note that the partition $\mathcal{P}_i$ of $\mathcal{S}^\circ$ does not only depend on the conic partition of $\mathcal{S}$, it also depends on the value of the matrices Q_i.

For instance, the cone defined by $2x+y \geq 0, x+2y \geq 0$ is the conic hull of the gradient of $\sqrt{x^2 - xy + y^2}$ evaluated at $(0,-1)$ and $(1,0)$ as shown in dotted arrows in Figure 5. The gradients are obtained by multiplying $(0,-1)$ and $(1,0)$ by the matrix Q_5. This illustrates why the partition of the polar is the image of the original partition under Q_5.

Example 4.6. For the piecewise semi-ellipsoidal template, the volume is not directly maximized. Instead, for each full-dimensional cone of the fan, we compute the sum s of the normalized rays and consider the polytope obtained by intersecting the cone with the halfspace $s^\top x \leq \|s\|_2^2$. We integrate the quadratic form corresponding to this cone, i.e. $h^2(\mathcal{S}, x)$ over the polytope. The sum of the integrals over each polytope is the objective function we use. This can be seen as the generalization of the sum of the squared length of the semi-axes of the polar of the ellipsoid.

The solution for different fans is illustrated in fig. 6. The maximal piecewise semi-ellipsoidal controlled invariant set with the fan $\mathcal{F}([-1,1]^2)$ has the following support function:

$$\delta^*(x|\mathcal{S})^2 = \begin{cases} (x_1 - x_2)^2 & \text{if } x_1x_2 \leq 0, \\ x_1^2 - x_1x_2/2 + x_2^2 & \text{if } x_1x_2 \geq 0. \end{cases}$$

For the cones $x_1x_2 \leq 0$, the semi-ellipsoid matches the square and the maximal controlled invariant set. For the cones $x_1x_2 \geq 0$, the semi-ellipsoid matches the optimal solution for the ellipsoid of maximal volume. This illustrates one key feature of piecewise semi-ellipsoidal sets, they can combine advantages of both polyhedra and ellipsoids. It can be polyhedral on the directions where the maximal controlled invariant set is polyhedral and be ellipsoidal on the directions where the maximal controlled invariant set is smooth or requires many halfspaces in its representation. We can use the polar of the polytope resulting from the first fixed point iteration to generate a refined fan. With this fan, the optimal solution for piecewise semi-ellipsoids matches the optimal solution to the generic set program. The support function is given by:

$$\delta^*(x|\mathcal{S})^2 = \begin{cases} (x_1 - x_2)^2 & \text{if } x_1x_2 \leq 0, \\ x_1^2 & \text{if } x_2(x_1 - 2x_2) \geq 0, \\ (x_1/2 + x_2)^2 & \text{if } x_1(2x_2 - x_1) \geq 0. \end{cases}.$$

4.4 Basic semialgebraic sets

Basic semialgebraic sets are sets defined by finitely many polynomial inequalities. For degree one polynomials, this corresponds to polyhedra. For simplicity, in this section, we focus on the case with only one polynomial inequality. More precisely, we consider the sets that are the sublevel set of a nonnegative *homogeneous*[2] polynomial of degree $2d$ for some positive integer d. We refer to these sets as *polysets of degree* $2d$ and denote it by

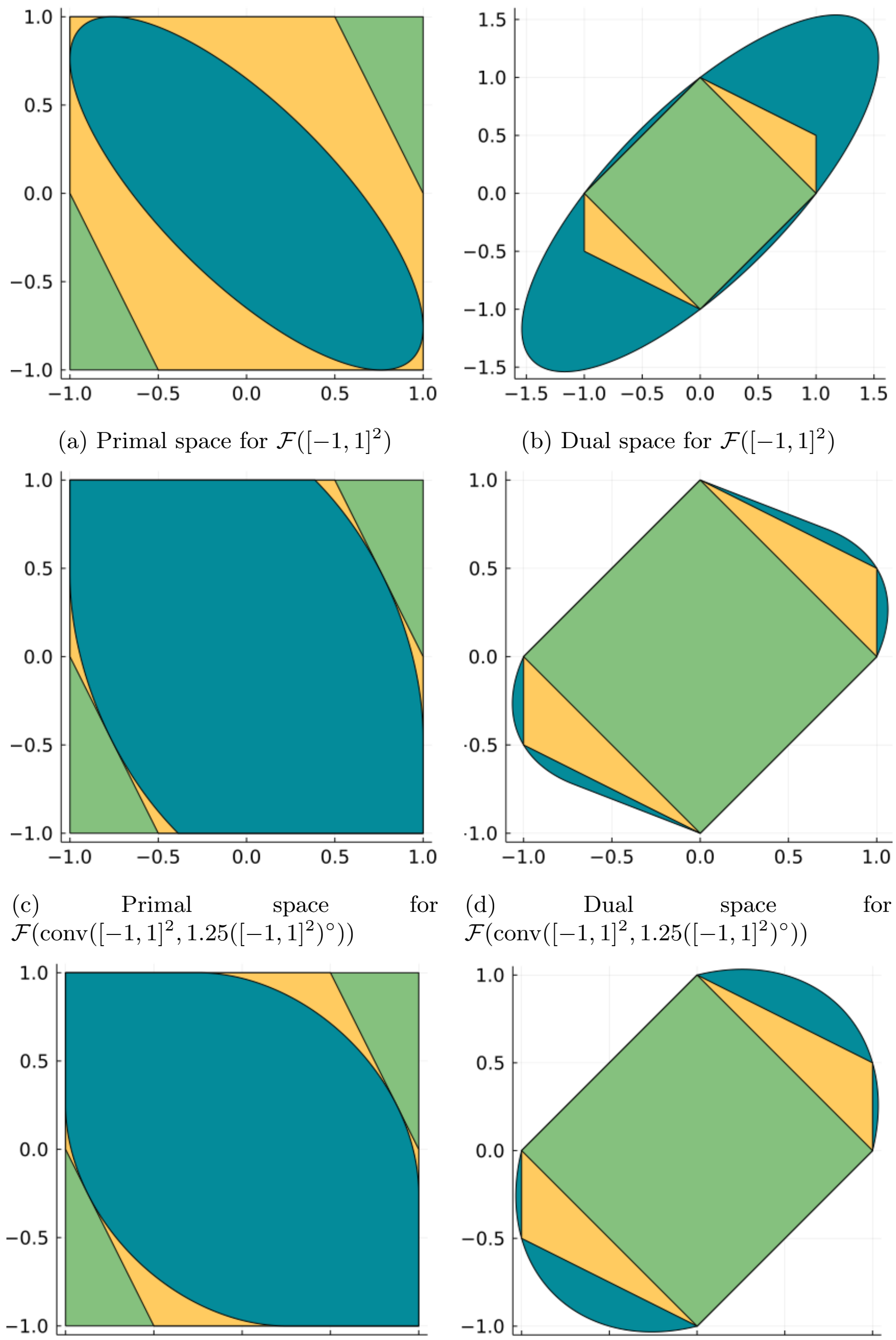
$$\mathcal{P}_p = \{\, x \in \mathbb{R}^n \mid p(x) \leq 1 \,\}. \tag{45}$$

where p is a homogeneous polynomial of degree $2d$. The family of ellipsoids presented in Section 4.2 is exactly the family of polysets of degree 2. We detail in this section how to generalize the results on ellipsoids to polysets of higher degree. Note that $p(x)$ does not have to be convex in eq. (45), and thus polysets can be non-convex as illustrated in the following example.

Example 4.7. Consider the discrete-time linear switched system (11) introduced in [12]:

$$f_1(x) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} x \qquad f_2(x) = \alpha \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} x$$

with $\alpha \approx 0.749$. Sets proving the best upper bound for the so-called *Joint Spectral Radius* (JSR) of the system are given in fig. 7. Although it is known that the JSR can be approximated arbitrarily closely by convex sets [24, Theorem 2.2]. For a polyset of a given degree, the possibility of polysets to be nonconvex allows to obtain tighter bounds on the JSR.



(a) Primal space for $\mathcal{F}([-1,1]^2)$

(b) Dual space for $\mathcal{F}([-1,1]^2)$

(c) Primal space for $\mathcal{F}(\mathrm{conv}([-1,1]^2, 1.25([-1,1]^2)^\circ))$

(d) Dual space for $\mathcal{F}(\mathrm{conv}([-1,1]^2, 1.25([-1,1]^2)^\circ))$

(e) Primal space for $\mathcal{F}(([-1,1]^2)^\circ)$

(f) Dual space for $\mathcal{F}(([-1,1]^2)^\circ)$

Figure 6: Optimal solutions for the set program of Example 4.1 for piecewise semi-ellipsoids. The green set is the square $[-1,1]^2$ and the blue and yellow sets are the optimal solutions with the objective replaced by the integral of $\delta^*(y|\mathcal{S})^2$ over the polar of the square and with different fans.

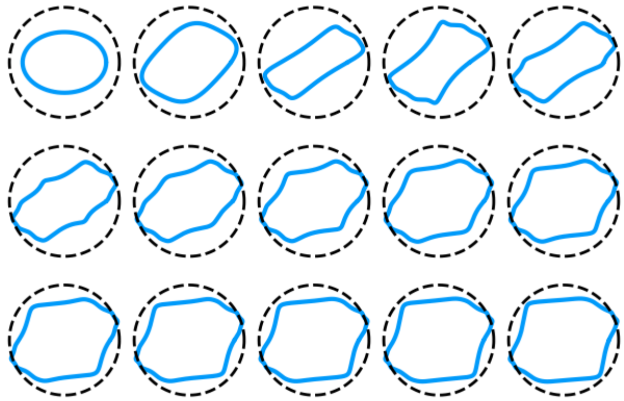


Figure 7: Sets providing the best upper bound to the JSR for Example 4.7. From left to right, the first row corresponds to $d = 1, 2, \dots, 5$, the second row corresponds to $d = 6, 7, \dots, 10$ and the third row corresponds to $d = 11, 12, \dots, 15$.

The gauge function and gauge-like function of degree $2d$ (see definition A.2) of the polyset are

$$\mathrm{g}(\mathcal{P}_p, x) = \sqrt[2d]{p(x)} \qquad\qquad \mathrm{g}_{2d}(\mathcal{P}_p, x) = \frac{p(x)}{2d}. \tag{46}$$

As a consequence of proposition A.9 and (46), the inclusion between polysets is verified using the following property.

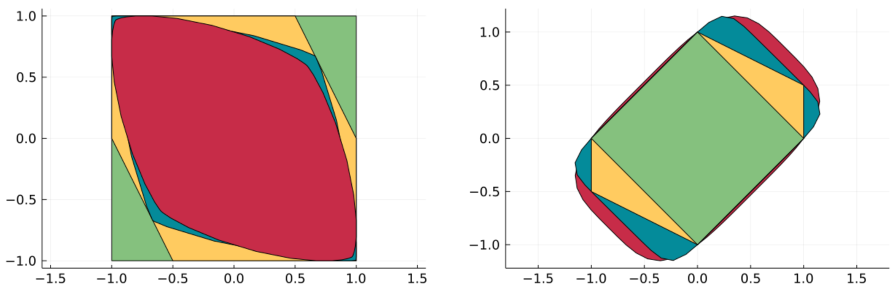
Proposition 4.2. Consider two polysets $\mathcal{P}_{p_1}, \mathcal{P}_{p_2}$ of degree $2d$. The inclusion $\mathcal{P}_{p_1} \subseteq \mathcal{P}_{p_2}$ is equivalent to $\forall x \in \mathbb{R}^n$, $p_1(x) \geq p_2(x)$.

Remark 4.3. Proposition 4.2 reformultes the inclusion as the nonnegativity of the polynomial $p_1(x) - p_2(x)$. Deciding the nonnegativity of a polynomial is however co-NP-hard. To render the problem tractable, the nonnegativity constraint is commonly restricted to a Sum-of-Squares constraint which can be reformulated as a semidefinite constraint [39]. Not all nonnegative polynomials are Sum-of-Squares but all Sum-of-Squares polynomials are nonnegative. Therefore, this restriction may discard some solutions but all solutions to the Sum-of-Squares restriction are also feasible to to the original program.

Well-known examples of polysets are the $(2d)$-unit balls. The gauge function of the $(2d)$-unit ball is the L^{2d} norm. The L^p norm is convex for any $p \geq 1$ and by proposition A.7 its polar is the q-unit ball for $q \geq 1$ such that $1/p + 1/q = 1$. Hence the $(2d)$-unit ball is convex for any positive integer d and its polar is the $(2d)/(2d-1)$-unit ball. This shows that, except for $d = 1$, the polar of a polyset is not necessarily a polyset (even if it is always a semialgebraic set).

Example 4.8. For the polyset template of degree $2d$, the dual representation is given by $\delta^*(y|\mathcal{S}) = \sqrt[2d]{p(x)}$ where $p(x)$ is a homogeneous polynomial of degree $2d$. We use the volume heuristic consisting in maximizing the integral of $p(x)$ over the square $[-1, 1]^2$. This is the method discussed in [15]. Note that, as discussed in [15, Remark 2], their heuristic coincides with the sum of squares of semi-axes for quadratic forms. For this reason, the optimal controlled invariant sublevel set of quadratic forms is the red ellipsoid in Figure 4. The optimal solution is given in Figure 8.

[2] A polynomial is homogeneous if all its monomials have the same total degree.



(a) Primal space (b) Dual space

Figure 8: Optimal solutions for the set program of Example 4.1 for polysets. The green set is the safe set $[-1,1]^2$, the yellow set is the optimal solution to the generic set program, the optimal solution to the set program for polysets of degree 4, (resp. 8, 18 and 22) is represented in orange (resp. blue, green and red).

5 Comparison of performance for convex bodies

Extensive research [4, 11, 40, 24, 42, 30] have resulted into approximation guarantees for ellipsoids and polysets for the computation of invariant sets of discrete-time switching linear systems, as defined in proposition 2.6. In this section, we generalize these results for a wider class of set programs.

Consider a set program of the form:

$$\begin{aligned} \underset{\mathcal{S}_1 \subseteq \mathbb{R}^{n_1}, \ldots, \mathcal{S}_N \subseteq \mathbb{R}^{n_N}}{\text{minimize}} \quad & \gamma \\ \text{subject to:} \quad & g_i(\mathcal{S}_1, \ldots, \mathcal{S}_N) \subseteq \gamma h_i(\mathcal{S}_1, \ldots, \mathcal{S}_N), \quad i = 1, 2, \ldots, M \\ & \mathcal{S}_i \text{ is a symmetric convex body.} \end{aligned} \tag{47}$$

For this set program, a sort of "measure of conservativeness" is given by the ratio of the optimal objective value between the generic set program and the set program for a given template. For the ellipsoidal template, theorem 5.2 provides a guaranteed bound on this ratio depending only on the dimensions of the spaces, i.e., $n_1, \ldots, n_N$. This bound relies on the *John's Ellipsoid Theorem*.

Theorem 5.1 (John's Ellipsoid Theorem [23])**.** For any symmetric convex body $\mathcal{S} \subseteq \mathbb{R}^n$, there exists a symmetric positive semidefinite matrix $Q \subseteq \mathbb{R}^{n \times n}$ such that

$$\mathcal{E}_Q \subseteq \mathcal{S} \subseteq \sqrt{n}\mathcal{E}_Q$$

We say that a set-valued function f is *strictly positively homogeneous* if for all sets $\mathcal{S}_1 \subseteq \mathbb{R}^{n_1}, \ldots, \mathcal{S}_N \subseteq \mathbb{R}^{n_N}$ and all positive real number $\gamma > 0$, we have $f(\gamma\mathcal{S}_1, \ldots, \gamma\mathcal{S}_N) = \gamma f(\mathcal{S}_1, \ldots, \mathcal{S}_N)$.

Theorem 5.2. Consider Program (47), let $\gamma^\star$ be the optimal objective value of the generic set program and $\gamma^\star_{\text{ell}}$ be the optimal objective value for the ellipsoidal template. If $\gamma^\star$ is finite, the functions h_i, g_i are monotone and the functions h_i are strictly positively homogeneous then

$$\gamma^\star \le \gamma^\star_{\text{ell}} \le \gamma^\star \max_{i \in [n]} \sqrt{n_i}.$$

Proof. The inequality $\gamma^\star \le \gamma^\star_{\text{ell}}$ comes from the fact that $\gamma^\star_{\text{ell}}$ is the optimal objective value of a restriction to a given template. Let $\mathcal{S}_i^\star$ be an optimal solution to the generic set program. By theorem 5.1, there exists a symmetric positive semidefinite matrix $Q_i \in \mathbb{R}^{n_i \times n_i}$ such that $\mathcal{E}_{Q_i} \subseteq \mathcal{S}_i^\star \subseteq \sqrt{n_i}\mathcal{E}_{Q_i}$. By the monotonicity of g_i, h_i we have

$$\begin{aligned} g_i(\mathcal{E}_{Q_1}, \ldots, \mathcal{E}_{Q_N}) &\subseteq g_i(\mathcal{S}_i^\star, \ldots, \mathcal{S}_N^\star) \\ &\subseteq \gamma^\star h_i(\mathcal{S}_i^\star, \ldots, \mathcal{S}_N^\star) \\ &\subseteq \gamma^\star h_i(\sqrt{n_1}\mathcal{E}_{Q_1}, \ldots, \sqrt{n_N}\mathcal{E}_{Q_N}) \\ &\subseteq \gamma^\star h_i(\max_{i \in [n]} \sqrt{n_i}\mathcal{E}_{Q_1}, \ldots, \max_{i \in [n]} \sqrt{n_i}\mathcal{E}_{Q_N}) \end{aligned}$$

We can conclude by the homogeneity of h_i that $(\mathcal{E}_{Q_i})_{i=1}^N$, $\gamma^\star \max_{i\in[n]} \sqrt{n_i}$ is a feasible solution to the set program for the ellipsoidal template. $\square$

Theorem 5.1 can be generalized for polysets as follows:

Theorem 5.3 ([8, Theorem 3.4])**.** For any symmetric convex body $\mathcal{S} \subseteq \mathbb{R}^n$, there exists a Sum-of-Squares polynomial p such that

$$\mathcal{P}_p \subseteq \mathcal{S} \subseteq \binom{n+d-1}{n}^{\frac{1}{2n}} \mathcal{P}_p.$$

Theorem 5.3 allows to generalize theorem 5.2 for polysets as follows. The proof is omitted for conciseness as it can be directly obtained from the proof of theorem 5.2 by replacing ellipsoids by polysets of degree $2d$ and theorem 5.1 by theorem 5.3.

Theorem 5.4. Consider Program (47), let $\gamma^\star$ be the optimal objective value of the generic set program and $\gamma^\star_{2d}$ be the optimal objective value for the template of polysets of degree $2d$. If $\gamma^\star$ is finite and h_i, g_i are monotone then

$$\gamma^\star \le \gamma^\star_{2d} \le \gamma^\star \max_{i\in[n]} \binom{n+d-1}{n}^{1/2d}.$$

The guarantees given by theorem 5.2 and theorem 5.4 are independent of the particular structure formed by the constraints of the set program. In the rest of this section, we describe an alternative guarantee that relies on this structure but is independent of the dimensions of the sets. This result is restricted to a particular case of g_i, h_i and we represent the constraint structure with a graph.

Consider m matrices $A_1, \ldots, A_m \in \mathbb{R}^{n\times n}$ and a directed and labelled strongly connected graph $G(V, E)$ with nodes V, edges E and possibly with parallel edges such that no node has zero ingoing or outgoing degree. Consider the following set program:

$$\begin{aligned} \underset{\mathcal{S}_1\subseteq\mathbb{R}^{n_1},\ldots,\mathcal{S}_N\subseteq\mathbb{R}^{n_N}}{\text{minimize}} \quad & \gamma \\ \text{subject to:} \quad A_\sigma \mathcal{S}_i & \subseteq \gamma \mathcal{S}_j, \quad \forall (i,j,\sigma) \in E \\ & \mathcal{S}_i \text{ is a symmetric convex body.} \end{aligned} \tag{48}$$

The optimal value for the generic set program is called the *constrained joint spectral radius* of the corresponding switched system [16].

Let E_k be the subset of E^k containing all valid paths of length k in the graph $G(V, E)$. We define the the *entropy of the edge shift* of the language of labels of paths of $G(V, E)$ [35, Definition 2.2.5, Definition 4.1.1] as:

$$h(E) = \lim_{k\to\infty} \frac{1}{k} \log_2 |E_k|. \tag{49}$$

Theorem 5.5 ([30])**.** Consider Program (48), let $\gamma^\star$ be the optimal objective value of the generic set program and $\gamma^\star_{2d}$ be the optimal objective value for the template of polysets of degree $2d$. If $\gamma^\star$ is finite then

$$\gamma^\star \le \gamma^\star_{2d} \le \gamma^\star 2^{\frac{h(E)}{2d}}.$$

Remark 5.1. The guarantees provided by theorem 5.4 and theorem 5.5 are quite appealing as they ensures that the conservativeness of the polyset template vanishes when the degree d tends to infinity.

There is however a gap between the statement of the theorems and the polyset solutions that can be computed numerically. Indeed, whether the sets $\mathcal{S}_i$ are represented using their gauge or support function, the inclusion

$$g_i(\mathcal{S}_1, \ldots, \mathcal{S}_N) \subseteq \gamma h_i(\mathcal{S}_1, \ldots, \mathcal{S}_N)$$

is reformulated as a polynomial nonnegativity which is restricted to a Sum-of-Squares constraint for numerical computation as discussed in remark 4.3. As the optimal solution to the set program for the polyset template might be represented by a nonnegative polynomial that is not a Sum-of-Squares. That is, the feasible polyset solutions that can be computed in practice with Sum-of-Squares programming may not have the same guarantee.

Nevertheless, for the particular case of Program (48), [42, Theorem 3.4] (resp. [30]) showed that theorem 5.4 (resp. theorem 5.5) still holds with the Sum-of-Squares restriction; the case of $N = |V| = 1$ was shown earlier in [40, Theorem 3.4] for both theorems.

6 Conclusion

The numerical computation of solutions to set programs is central in Dynamics and Control. The rich diversity of dynamical systems and control problems translates into a variety of combinations of inclusion constraints. Different templates have fundamentally different and sometimes complementary computational and numerical behaviors that vary considerably depending on the class of set programs.

This motivates the need for studying various different templates and adapt the selection depending on the particular class or even instance of set program. Fortunately, as detailed in the chapter, the choice of representation and the reformulation of the inclusion as functional inequality is template-independent which enables the analysis of set programs in a template-independent manner as shown section 3.

Numerous control problem can be tackled by combining the techniques developed in this chapter. For conciseness, the invariance related to continuous-time dynamics were omitted. A systematic analysis of the inclusion constraints related to these invariance conditions, combined with the discrete-time results of this chapter provides a complete procedure for computing invariant sets for hybrid systems [28].

There is an unlimited number of possible inclusion constraints that can be obtained by combining the operations of table 1, let alone other operations such as set complements. This chapter covers the most common ones and noteworthy other forms of inclusion constraints remain to be investigated such as eq. (21) with nonlinear functions.

The performance comparison based on the John's ellipsoid theorem covers a wide range of set programs. More templates remain to be added to the analysis such as piecewise semi-ellipsoids or ellipsoids with different centers. Moreover, since the bound of theorem 5.5 is better than the bound of theorem 5.4 for a sufficiently large degree $2d$ of the polyset template, it would be valuable to enlarge the class of set programs for which this bound applies.

References

[1] Amir Ali Ahmadi and Raphael M Jungers. "Switched stability of nonlinear systems via sos-convex lyapunov functions and semidefinite programming". In: *2013 IEEE 52nd Annual Conference on Decision and Control (CDC)*. IEEE. 2013, pp. 727–732.

[2] Amir Ali Ahmadi et al. "Joint spectral radius and path-complete graph Lyapunov functions". In: *SIAM Journal on Control and Optimization* 52.1 (2014), pp. 687–717. DOI: 10.1137/110855272.

[3] Matthias Althoff, Goran Frehse, and Antoine Girard. "Set Propagation Techniques for Reachability Analysis". In: *Annual Review of Control, Robotics, and Autonomous Systems* 4.1 (May 2021). DOI: 10.1146/annurev-control-071420-081941.

[4] Tsuyoshi Ando and Mau-hsiang Shih. "Simultaneous Contractibility." In: *SIAM Journal on Matrix Analysis & Applications* 19.2 (1998), p. 487. ISSN: 08954798. DOI: 10.1137/S0895479897318812.

[5] David Avis. "A revised implementation of the reverse search vertex enumeration algorithm". In: *Polytopes — combinatorics and computation*. Springer. 2000, pp. 177–198. DOI: 10.1007/978-3-0348-8438-9_9.

[6] David Avis, David Bremner, and Raimund Seidel. "How good are convex hull algorithms?" In: *Proceedings of the eleventh annual symposium on Computational geometry*. ACM. 1995, pp. 20–28. DOI: 10.1016/S0925-7721(96)00023-5.

[7] Francis Bach. “Submodular functions: from discrete to continuous domains”. In: *Mathematical Programming* 175.1-2 (2019), pp. 419–459. DOI: 10.1007/s10107-018-1248-6.

[8] Alexander Barvinok. *A course in convexity*. Vol. 54. American Mathematical Society Providence, 2002. DOI: 10.1090/gsm/054.

[9] Andrew An Bian et al. “Guaranteed non-convex optimization: Submodular maximization over continuous domains”. In: *arXiv preprint arXiv:1606.05615* (2016).

[10] Franco Blanchini and Stefano Miani. *Set-theoretic methods in control*. Second. Springer, 2015. DOI: 10.1007/978-0-8176-4606-6.

[11] Vincent D Blondel and Yurii Nesterov. “Computationally efficient approximations of the joint spectral radius”. In: *SIAM Journal on Matrix Analysis and Applications* 27.1 (2005), pp. 256–272. DOI: 10.1137/040607009.

[12] Vincent D Blondel, Jacques Theys, and Alexander A Vladimirov. “An elementary counterexample to the finiteness conjecture”. In: *SIAM Journal on Matrix Analysis and Applications* 24.4 (2003), pp. 963–970.

[13] Vincent D Blondel and John N Tsitsiklis. “The boundedness of all products of a pair of matrices is undecidable”. In: *Systems & Control Letters* 41.2 (2000), pp. 135–140. DOI: 10.1016/S0167-6911(00)00049-9.

[14] Brian Borchers. “CSDP, A C library for semidefinite programming”. In: *Optimization methods and Software* 11.1-4 (1999), pp. 613–623. DOI: 10.1080/10556789908805765.

[15] Fabrizio Dabbene, Didier Henrion, and Constantino M Lagoa. “Simple approximations of semialgebraic sets and their applications to control”. In: *Automatica* 78 (2017), pp. 110–118. DOI: 10.1016/j.automatica.2016.11.021.

[16] Xiongping Dai. “A Gel’fand-type spectral radius formula and stability of linear constrained switching systems”. In: *Linear Algebra and its Applications* 436.5 (2012), pp. 1099–1113. DOI: 10.1016/j.laa.2011.07.029.

[17] Iain Dunning, Joey Huchette, and Miles Lubin. “JuMP: A modeling language for mathematical optimization”. In: *SIAM Review* 59.2 (2017), pp. 295–320. DOI: 10.1137/15M1020575.

[18] K Fukuda. “Note on New Complexity Classes ENP, EP and CEP—an Extension of the Classes NP Co-NP and P”. In: *ETH Zürich, Institute for Operations Research, Zürich June* 12 (1996), p. 1996.

[19] Komei Fukuda. “CDD–A C-implementation of the double description method”. In: *Institut fur Operations Research, ETH Zurich* (2003).

[20] Rafal Goebel, Ricardo G Sanfelice, and Andrew R Teel. “Hybrid dynamical systems”. In: *IEEE Control Systems* 29.2 (2009), pp. 28–93. DOI: 10.1515/9781400842636.

[21] Rafal Goebel et al. “Conjugate convex Lyapunov functions for dual linear differential inclusions”. In: *Automatic Control, IEEE Transactions on* 51.4 (2006), pp. 661–666. DOI: 10.1109/TAC.2006.872764.

[22] Didier Henrion and Milan Korda. “Convex computation of the region of attraction of polynomial control systems”. In: *IEEE Transactions on Automatic Control* 59.2 (2014), pp. 297–312. DOI: 10.1109/TAC.2013.2283095.

[23] Fritz John. “Extremum Problems with Inequalities as Subsidiary Conditions”. English. In: *Traces and Emergence of Nonlinear Programming*. Ed. by Giorgio Giorgi and Tinne Hoff Kjeldsen. First published in 1948. Springer Basel, 2014, pp. 197–215. ISBN: 978-3-0348-0438-7. DOI: 10.1007/978-3-0348-0439-4_9.

[24] Raphaël Jungers. *The joint spectral radius: theory and applications*. Vol. 385. Springer Science & Business Media, 2009. DOI: 10.1007/978-3-540-95980-9.

[25] Niklas Kochdumper and Matthias Althoff. “Constrained Polynomial Zonotopes”. In: *arXiv preprint arXiv:2005.08849* (2020).

[26] Milan Korda, Didier Henrion, and Colin N Jones. "Convex computation of the maximum controlled invariant set for polynomial control systems". In: *SIAM Journal on Control and Optimization* 52.5 (2014), pp. 2944–2969.

[27] Benoît Legat and Raphaël M. Jungers. *Computation of invariant sets for complex systems.* May 2021. DOI: 10.24433/CO.5065309.V1.

[28] Benoît Legat and Raphaël M. Jungers. "Geometric control of algebraic systems". In: *Proceedings of the 7th IFAC Conference on Analysis and Design of Hybrid Systems.* Vol. 54. 5. 7th IFAC Conference on Analysis and Design of Hybrid Systems ADHS 2021. 2021, pp. 79–84. DOI: 10.1016/j.ifacol.2021.08.478.

[29] Benoît Legat and Raphaël M. Jungers. *Geometric control of hybrid systems.* IFAC Journal of Nonlinear Analysis: Hybrid Systems. 2022.

[30] Benoît Legat, Pablo A. Parrilo, and Raphaël M. Jungers. "An Entropy-Based Bound for the Computational Complexity of a Switched System". In: *IEEE Transactions on Automatic Control* 64.11 (Nov. 2019), pp. 4623–4628. DOI: 10.1109/tac.2019.2902625.

[31] Benoît Legat, Saša V. Raković, and Raphaël M. Jungers. "Piecewise Semi-Ellipsoidal Control Invariant Sets". In: *IEEE Control Systems Letters* 5.3 (July 2021), pp. 755–760. DOI: 10.1109/LCSYS.2020.3005326.

[32] Benoît Legat et al. "MathOptInterface: a data structure for mathematical optimization problems". In: *INFORMS Journal on Computing* (2021). DOI: 10.1287/ijoc.2021.1067.

[33] Daniel Liberzon. *Switching in systems and control.* Springer Science & Business Media, 2012. DOI: 10.1007/978-1-4612-0017-8.

[34] Daniel Liberzon and Stephan Trenn. "Switched nonlinear differential algebraic equations: Solution theory, Lyapunov functions, and stability". In: *Automatica* 48.5 (2012), pp. 954–963. DOI: 10.1016/j.automatica.2012.02.041.

[35] Douglas Lind and Brian Marcus. *An introduction to symbolic dynamics and coding.* Cambridge university press, 1995. DOI: 10.1017/CBO9780511626302.

[36] László Lovász. "Submodular functions and convexity". In: *Mathematical Programming The State of the Art.* Springer, 1983, pp. 235–257. DOI: 10.1007/978-3-642-68874-4_10.

[37] Aleksandr Mikhailovich Lyapunov. "The general problem of the stability of motion". In: *International Journal of Control* 55.3 (1992), pp. 531–534. DOI: 10.1080/00207179208934253.

[38] David H Owens and Dragutin Lj Debeljkovic. "Consistency and Liapunov stability of linear descriptor systems: A geometric analysis". In: *IMA Journal of Mathematical Control and Information* 2.2 (1985), pp. 139–151. DOI: 10.1093/imamci/2.2.139.

[39] Pablo A Parrilo. "Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization". PhD thesis. Citeseer, 2000.

[40] Pablo A Parrilo and Ali Jadbabaie. "Approximation of the joint spectral radius using sum of squares". In: *Linear Algebra and its Applications* 428.10 (2008), pp. 2385–2402. DOI: 10.1016/j.laa.2007.12.027.

[41] Matthew Philippe et al. "On path-complete Lyapunov functions: Geometry and comparison". In: *IEEE Transactions on Automatic Control* 64.5 (2018), pp. 1947–1957.

[42] Matthew Philippe et al. "Stability of discrete-time switching systems with constrained switching sequences". In: *Automatica* 72 (2016), pp. 242–250. DOI: 10.1016/j.automatica.2016.05.015.

[43] Saša V. Raković. "Control Minkowski–Lyapunov functions". In: *Automatica* 128 (June 2021), p. 109598. DOI: 10.1016/j.automatica.2021.109598.

[44] Saša V. Raković. "Robust control Minkowski–Lyapunov functions". In: *Automatica* 125 (Mar. 2021), p. 109437. DOI: 10.1016/j.automatica.2020.109437.

[45] R Tyrrell Rockafellar and Roger JB Wets. "Variational Analysis". In: 317 (1998). ISSN: 0072-7830. DOI: 10.1007/978-3-642-02431-3.

[46] Ralph Tyrell Rockafellar. *Convex analysis*. Princeton university press, 2015. DOI: 10.1515/9781400873173.

[47] Bertrand Russell. *Principles of Mathematics*. Routledge, 1937.

[48] Rolf Schneider. *Convex bodies: the Brunn–Minkowski theory*. 151. Cambridge University Press, 2013. DOI: 10.1017/CBO9781139003858.

[49] Tillmann Weisser et al. "Polynomial and Moment Optimization in Julia and JuMP". In: *JuliaCon*. 2019.

[50] Günter M Ziegler. *Lectures on polytopes*. Vol. 152. Springer Science & Business Media, 1995. DOI: 10.1007/978-1-4613-8431-1.

A Convex sets: geometry and computation

A.1 Properties of set-valued maps

Consider a vector-valued function $f : \mathbb{R}^n \to \mathbb{R}^r$. The preimage f^{-1}, defined in section 1.1, is its right inverse:

$$\forall \mathcal{S} \subseteq \mathbb{R}^r, \quad f(f^{-1}(\mathcal{S})) = \mathcal{S} \tag{50}$$

On the other hand, if f is not injective, the preimage f^{-1} is not necessarily its left inverse but it still satisfies the following property:

$$\forall \mathcal{S} \subseteq \mathbb{R}^n, \quad f^{-1}(f(\mathcal{S})) \supseteq \mathcal{S}. \tag{51}$$

We say that a real-valued function f (resp. set-valued function h) is *monotone* if for all sets $\mathcal{S}_1 \subseteq \mathbb{R}^{n_1}, \ldots, \mathcal{S}_N \subseteq \mathbb{R}^{n_N}$ and $\mathcal{S}_1' \subseteq \mathbb{R}^{n_1}, \ldots, \mathcal{S}_N' \subseteq \mathbb{R}^{n_N}$ such that $\mathcal{S}_i \subseteq \mathcal{S}_i'$ for all $i \in [N]$, we have $f(\mathcal{S}_1, \ldots, \mathcal{S}_N) \le f(\mathcal{S}_1', \ldots, \mathcal{S}_N')$ (resp. $h(\mathcal{S}_1, \ldots, \mathcal{S}_N) \subseteq h(\mathcal{S}_1', \ldots, \mathcal{S}_N')$). If, in addition, $h(\mathcal{S}_1, \ldots, \mathcal{S}_N) \subseteq h(\mathcal{S}_1', \ldots, \mathcal{S}_N')$ implies that $\mathcal{S}_i \subseteq \mathcal{S}_i'$ then h is an *order embedding*.

Given a vector-valued map, the set-valued function mapping sets to their images (resp. preimages) is monotone (resp. an order embedding). That is,

$$\forall \mathcal{S}, \mathcal{T} \subseteq \mathbb{R}^n, \qquad \mathcal{S} \subseteq \mathcal{T} \Rightarrow f(\mathcal{S}) \subseteq f(\mathcal{T}), \tag{52}$$

$$\forall \mathcal{S}, \mathcal{T} \subseteq \mathbb{R}^r, \qquad \mathcal{S} \subseteq \mathcal{T} \Leftrightarrow f^{-1}(\mathcal{S}) \subseteq f^{-1}(\mathcal{T}). \tag{53}$$

When working with inclusion constraints, it is tempting to rewrite an image (resp. preimage) from one side of the inclusion as a preimage (resp. image) to the other side. As we show in proposition A.1 and remark A.1, this is only possible in specific cases.

Proposition A.1. Consider a function $f \in \mathbb{R}^n \to \mathbb{R}^r$, a set $\mathcal{S} \subseteq \mathbb{R}^n$ and a set $\mathcal{T} \subseteq \mathbb{R}^r$. The inclusion $f(\mathcal{S}) \subseteq \mathcal{T}$ holds if and only if the inclusion $\mathcal{S} \subseteq f^{-1}(\mathcal{T})$ holds.

Proof. We can prove the proposition by combining the relations (50), (51), (53) and (52) as follows:

$$\begin{aligned} f(\mathcal{S}) \subseteq \mathcal{T} &\overset{(53)}{\Leftrightarrow} f^{-1}(f(\mathcal{S})) \subseteq f^{-1}(\mathcal{T}) \overset{(51)}{\Rightarrow} \mathcal{S} \subseteq f^{-1}(\mathcal{T}) \\ f(\mathcal{S}) \subseteq \mathcal{T} &\overset{(50)}{\Leftrightarrow} \quad f(\mathcal{S}) \subseteq f(f^{-1}(\mathcal{T})) \overset{(52)}{\Leftarrow} \mathcal{S} \subseteq f^{-1}(\mathcal{T}). \end{aligned}$$

□

Remark A.1. Note that $\mathcal{S} \subseteq f(\mathcal{T})$ does not imply $f^{-1}(\mathcal{S}) \subseteq \mathcal{T}$. For instance, if $A = \begin{bmatrix} 1 & 0 \end{bmatrix}$, $\mathcal{S} = [-1, 1]$ and $\mathcal{T} = [-1, 1]^2$ then $\mathcal{S} = [-1, 1] = A\mathcal{T}$ but $A^{-1}\mathcal{S} = [-1, 1] \times \mathbb{R} \not\subseteq \mathcal{T}$.

Proposition A.2. Given a number $\alpha \in \mathbb{R}$, functions $g : \mathbb{R}^{n_1} \to \mathbb{R}$ and $f : \mathbb{R}^{n_2} \to \mathbb{R}^{n_1}$, the following holds:

$$f^{-1}(\{\, x \mid g(x) \le \alpha \,\}) = \{\, y \mid g(f(y)) \le \alpha \,\}. \tag{54}$$

Proof. Given any $y \in \mathbb{R}^{n_2}$, $y \in f^{-1}(\{\, x \mid g(x) \le \alpha \,\})$ if and only if $f(y) \in \{\, x \mid g(x) \le \alpha \,\}$ which is equivalent to $g(f(y)) \le \alpha$. □

Definition A.1. A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be *positively homogeneous* (of degree 1) if for every $x \in \mathbb{R}^n$ and real number $\lambda > 0$, we have $f(\lambda x) = \lambda f(x)$.

Proposition A.3. Given a positive real number $\alpha > 0$, positively homogeneous functions $g_i : \mathbb{R}^n \to \mathbb{R}_+$, and the sets $\mathcal{S}_i = \{\, x \mid g_i(x) \le \alpha \,\}$ for $i = 1, 2$. The inclusion

$$\mathcal{S}_1 \subseteq \mathcal{S}_2 \tag{55}$$

is equivalent to

$$\forall x \in \mathbb{R}^n, \quad g_1(x) \ge g_2(x). \tag{56}$$

Proof. Suppose by contradiction that there exists $x \in \mathcal{S}_1 \setminus \mathcal{S}_2$. This means that $g_1(x) \leq \alpha < g_2(x)$ which contradicts eq. (56). Suppose by contradiction that there exists x such that $g_1(x) < g_2(x)$. If $g_1(x) = 0$ (resp. $g_1(x) > 0$), let $y = (\alpha + 1)x/g_2(x)$ (resp. $y = \alpha x/g_1(x)$) then by positive homogeneity we have $g_1(y) = 0 < \alpha < g_2(y)$ (resp. $\alpha = g_1(y) < g_2(y)$) hence $y \in \mathcal{S}_1 \setminus \mathcal{S}_2$ which contradicts eq. (55). $\square$

A.2 Functional representation of convex sets

In this section, we discuss the relation between the different functions that represent a convex sets: the *gauge function*, the *indicator function* and the *support function*.

Definition A.2 (Minkowski function)**.** We define the *gauge* or *Minkowski* function $\mathrm{g}(\mathcal{S}, \cdot) : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ of a closed convex set $\mathcal{S}$ containing the origin as:

$$\mathrm{g}(\mathcal{S}, x) \triangleq \min_{\gamma} \{\, \gamma : x \in \gamma\mathcal{S}, \gamma \geq 0 \,\}, \text{ for } x \in \mathbb{R}^n.$$

The set $\mathcal{S}$ is the 1-sublevel set of its Minkowski function and the Minkowski function is the only positively homogeneous function that has this property.

Remark A.2. As we do not assume boundedness of the set in definition A.2, the Minkowski function can be zero for nonzeros vectors x. Moreover, as we do not assume that the set has nonempty interior and that the origin is in the interior of the set, the Minkowski function may be infinite for some vectors x. The image set of Minkowski functions is the set of extended reals for these reasons. As we compare Minkowski functions in this chapter, we adopt the convention $\infty = \infty$ so that the inequality $\mathrm{g}(\mathcal{S}_1, x) \geq \mathrm{g}(\mathcal{S}_2, x)$ holds when the value on both sides is ∞. See [46, p. 23] or [45, Example 3.50] for more details.

Definition A.3 (Indicator function)**.** We define the *indicator function* $\delta(x|\mathcal{S}) : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ of a set $\mathcal{S} \subseteq \mathbb{R}^n$ as:

$$\delta(x|\mathcal{S}) \triangleq \begin{cases} 0 & \text{if } x \in \mathcal{S} \\ +\infty & \text{otherwise.} \end{cases}$$

Definition A.4 (Conjugate function)**.** The *conjugate*, also called the *Legendre-Fenchel transformation*, of a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is the function

$$f^*(v) \triangleq \sup\{\, \langle v, x \rangle - f(x) \mid x \in \mathbb{R}^n \,\}, \text{ for } v \in \mathbb{R}^n.$$

The conjugate function of the indicator function is the *support function*.

Definition A.5 (Support function)**.** We define the *support function* $\delta^*(\cdot|\mathcal{S}) : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ of a nonempty closed convex subset $\mathcal{S} \subseteq \mathbb{R}^n$ as:

$$\delta^*(y|\mathcal{S}) \triangleq \sup_{x \in \mathcal{S}} \langle y, x \rangle.$$

Definition A.6 (Polar set [46, p. 125], [48, p. 32])**.** Given a subset $\mathcal{S} \subseteq \mathbb{R}^n$, the *polar set*, denoted $\mathcal{S}^\circ$, is defined as follows:

$$\mathcal{S}^\circ \triangleq \{\, y \in \mathbb{R}^n \mid \langle x, y \rangle \leq 1, \forall x \in \mathcal{S} \,\}.$$

The polar transformation satisfies the following property: For any convex sets $\mathcal{S}, \mathcal{T}$,

$$\mathcal{S} \subseteq \mathcal{T} \Leftrightarrow \mathcal{T}^\circ \subseteq \mathcal{S}^\circ. \tag{57}$$

Proposition A.4 ([46, Theorem 14.5], [48, Lemma 1.7.13])**.** Given a closed convex subset $\mathcal{S} \subseteq \mathbb{R}^n$ containing the origin, the following holds:

$$\forall x \in \mathbb{R}^n, \qquad \mathrm{g}(\mathcal{S}, x) = \delta^*(x|\mathcal{S}^\circ).$$

While proposition A.4 allows to go from the gauge function of a set to the support function of its polar, it is sometimes desirable to compute the support function of the set itself or the gauge function of its polar.

Definition A.7 (Polar function [46, p. 136]). The *polar* of a nonnegative convex function f such that $f(0) = 0$ is the function

$$f^\circ(v) \triangleq \inf\{\, \mu \geq 0 \mid \langle v, x\rangle \leq 1 + \mu f(x), \forall x \,\}. \tag{58}$$

Remark A.3. Note that if $f(x)$ is positively homogeneous, then eq. (58) is equivalent to

$$f^\circ(v) \triangleq \inf\{\, \mu \geq 0 \mid \langle v, x\rangle \leq \mu f(x), \forall x \,\}.$$

Proposition A.5 ([46, Corollary 15.1.2]). If $\mathcal{S}$ is a closed convex set containing the origin then the gauge function of $\mathcal{S}$ and the support function of $\mathcal{S}$ are polar to each other.

Proposition A.6 ([46, Corollary 15.3.2]). Let f be a positively homogeneous function of degree p, where $1 < p < \infty$. Then the polar of $(pf)^{1/p}$ is $(qf^*)^{1/q}$, where $1 < q < \infty$ and $1/p + 1/q = 1$.

Proposition A.7. The polar of the unit ball of the p-norm for $1 < p < \infty$ is the unit ball of the q-norm where $1 < q < \infty$ and $1/p + 1/q = 1$.

Proof. Consider the function $f(x) = (x_1^p + \cdots + x_n^p)/p$. The conjugate of $f(x)$ is $f^*(x) = (x_1^q + \cdots + x_n^q)/q$ where $1 < q < \infty$ is such that $1/p + 1/q = 1$. As $(pf)^{1/p}$ is the p-norm and $(qf^*)^{1/q}$ is the q-norm, we can conclude with proposition A.6. □

Given a closed convex set $\mathcal{S}$ containing the origin and a number $p \geq 1$, the function $\mathrm{g}^p(\mathcal{S}, x)/p$ is denoted $\mathrm{g}_p(\mathcal{S}, x)$ and referred to as the *gauge-like function of degree p*.

The inclusion of convex sets is equivalent to an inequality between the gauge functions.

Proposition A.8. Consider two closed convex subsets $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{R}^n$ containing the origin. The inclusion $\mathcal{S}_1 \subseteq \mathcal{S}_2$ is equivalent to the inequality $\mathrm{g}(\mathcal{S}_1, x) \geq \mathrm{g}(\mathcal{S}_2, x)$ for all $x \in \mathbb{R}^n$.

Proof. Follows from proposition A.3. □

This property can be generalized to gauge-like functions of arbitrary degree.

Proposition A.9. Consider two closed convex subsets $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{R}^n$ containing the origin and a positive integer p. The inclusion $\mathcal{S}_1 \subseteq \mathcal{S}_2$ is equivalent to the inequality $\mathrm{g}_p(\mathcal{S}_1, x) \geq \mathrm{g}_p(\mathcal{S}_2, x)$ for all $x \in \mathbb{R}^n$.

Proof. Follows from proposition A.3 and the fact that the function $x \mapsto x^p$ is strictly increasing for $x \in \mathbb{R}_+$ if $p \geq 1$, □

The inclusion can be rewritten as an inequality between the support functions as well.

Proposition A.10 ([46, Corollary 13.1.1]). Consider two nonempty closed convex subsets $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{R}^n$. The inclusion $\mathcal{S}_1 \subseteq \mathcal{S}_2$ is equivalent to the inequality $\delta^*(x|\mathcal{S}_1) \leq \delta^*(x|\mathcal{S}_2)$ for all $x \in \mathbb{R}^n$.

Definition A.8 (Proper convex function [46, p. 24]). We say that a convex function $f(x)$ is *proper* if its epigraph is non-empty and $f(x) < +\infty$ for at least one x and $f(x) > -\infty$ for every x.

Let $\partial f(x)$ denote the set of *subdifferential* of f at x. The conjugacy and subgradients are related by the following property.

Proposition A.11 ([46, Theorem 23.5]). For any proper convex function f and vectors x, y, $y \in \partial f(x)$ if and only if $x \in \partial f^*(y)$.

For differentiable functions, this allows to compute the conjugate of a function by inverting its gradient.

Corollary A.1. For any proper convex differentiable function f and vectors x, y, $y = \nabla f(x)$ if and only if $x = \nabla f^*(y)$.

	$\cap$	$\operatorname{conv}\cup$	$+$	$\sharp$	A	A^{-1}
Gauge function	max	$\square$	$\sharp$	$+$		$\circ A$
Support function	$\square$	max	$+$	$\sharp$	$\circ A^\top$	
Polar set	$\operatorname{conv}\cup$	$\cap$	$\sharp$	$+$	$A^{-\top}$	$A^\top$

Table 2: Given an operation on convex sets specified by the column header, the corresponding operation on the gauge function, support function or polar set is provided in the table. The symbols $\circ A$ denotes the composition with the linear map A. The symbol $\square$ (resp. $\sharp$) denotes the infimum convolution (resp. inverse sum) defined in eq. (59) (resp. eq. (60)). The two missing entries denote that there is no "direct link" with an operation on the function to the best of our knowledge, unless A is invertible.

A.3 Operations preserving convexity

We consider in this section six operations on sets that preserve the convexity of its operands: the intersection $\cap$, the convex hull of the union "$\operatorname{conv}\cup$", the Minkowski sum $+$, the inverse sum $\sharp$ (defined in eq. (60)), the linear image A and the linear preimage A^{-1}. In particular, we study their impact on the gauge function, support function and polar set. The results are summarized in table 2.

Lemma A.1 ([46, Corollary 16.5.2])**.** *Let $\mathcal{S}_i$ be a convex set in $\mathbb{R}^n$ for each $i \in I$. Then*

$$\Big(\operatorname{conv}\bigcup_{i\in I}\mathcal{S}_i\Big)^\circ = \bigcap_{i\in I}\mathcal{S}_i^\circ$$

$$(\bigcap_{i\in I}\operatorname{cl}\mathcal{S}_i)^\circ = \operatorname{cl}\Big(\operatorname{conv}\bigcup_{i\in I}\mathcal{S}_i^\circ\Big).$$

The gauge function of the intersection of sets is obtained as follows in terms of the support functions of the two sets being intersected.

Proposition A.12. *Given closed convex sets $\mathcal{S}_1, \mathcal{S}_2$ containing the origin, the following holds:*

$$\mathrm{g}(\mathcal{S}_1 \cap \mathcal{S}_2, x) = \max(\mathrm{g}(\mathcal{S}_1, x), \mathrm{g}(\mathcal{S}_2, x)).$$

The support function of the intersection of sets is the *infimal convolution* of the support function of each set.

Proposition A.13 (Infimal convolution [46, Theorem 5.4])**.** *Let $f_1, \ldots, f_m$ be proper convex functions, their infimal convolution*

$$(f_1 \,\square\, f_2 \,\square \cdots \square\, f_m)(x) \triangleq \inf\{\, f_1(x_1) + \cdots, f_m(x_m) \mid x_1 + \cdots + x_m = x \,\} \tag{59}$$

is convex.

Proposition A.14 ([46, Corollary 16.4.1])**.** *Given nonempty closed convex sets $\mathcal{S}_1, \mathcal{S}_2$, the following holds:*

$$\delta^*(y|\mathcal{S}_1 \cap \mathcal{S}_2) = \big[\delta^*(\cdot|\mathcal{S}_1) \,\square\, \delta^*(\cdot|\mathcal{S}_2)\big](y).$$

The support function of the convex hull of the union of sets is obtained as follows in terms of the support functions of the two sets.

Proposition A.15 ([46, Corollary 16.5.1])**.** *Given nonempty closed convex sets $\mathcal{S}_1, \mathcal{S}_2$, the following holds:*

$$\delta^*(y|\operatorname{conv}(\mathcal{S}_1 \cup \mathcal{S}_2)) = \max(\delta^*(y|\mathcal{S}_1), \delta^*(y|\mathcal{S}_2)).$$

Proof. It follows from lemma A.1, proposition A.4 and proposition A.12. $\square$

The gauge function of the convex hull is obtained in as an infimal convolution.

Proposition A.16. Given closed convex sets $\mathcal{S}_1, \mathcal{S}_2$ containing the origin, the following holds:

$$\mathrm{g}(\mathrm{conv}(\mathcal{S}_1 \cup \mathcal{S}_2), x) = \left[\, \mathrm{g}(\mathcal{S}_1, \cdot) \,\square\, \mathrm{g}(\mathcal{S}_2, \cdot)\right](y).$$

Proof. It follows from lemma A.1, proposition A.4 and proposition A.14. $\square$

The support function of the Minkowski sum of sets is obtained as follows in terms of the support functions of the two sets being summed.

Proposition A.17 ([46, Corollary 16.4.1])**.** Given nonempty closed convex sets $\mathcal{S}_1, \mathcal{S}_2$, the following holds:

$$\delta^*(y|\mathcal{S}_1 + \mathcal{S}_2) = \delta^*(y|\mathcal{S}_1) + \delta^*(y|\mathcal{S}_2).$$

The gauge function of the Minkowski sum of sets is the *inverse sum* of the gauge function of each set.

Proposition A.18 (Inverse sum [46, Theorem 5.8])**.** Let $f_1, \ldots, f_m$ be proper convex functions, their *inverse sum*

$$(f_1 \,\sharp\, f_2 \,\sharp \cdots \sharp\, f_m)(x) = \inf\{\, \max(f_1(x_1), \ldots, f_m(x_m)) \mid x_1, \ldots, x_m \in \mathbb{R}^m : x_1 + \cdots + x_m = x \,\}. \tag{60}$$

is convex.

Proposition A.19. Given closed convex sets $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{R}^n$ containing the origin, the following holds:

$$\mathrm{g}(\mathcal{S}_1 + \mathcal{S}_2, x) = \left[\, \mathrm{g}(\mathcal{S}_1, \cdot) \,\sharp\, \mathrm{g}(\mathcal{S}_2, \cdot)\right](x). \tag{61}$$

Proof. We have $x \in \mathcal{S}_1 + \mathcal{S}_2$ if and only if there exists $x_1 \in \mathcal{S}_1, x_2 \in \mathcal{S}_2$ such that $x_1 + x_2 = x$. That is, $\mathrm{g}(\mathcal{S}_1 + \mathcal{S}_2, x) \le 1$ if and only if there exists x_1, x_2 such that $\max(\mathrm{g}(\mathcal{S}_1, x_1), \mathrm{g}(\mathcal{S}_2, x_2)) \le 1$ and $x_1 + x_2 = x$. In other words, we have

$$\mathrm{g}(\mathcal{S}_1 + \mathcal{S}_2, x) \le 1 \qquad \Leftrightarrow \qquad \inf\{\, \max(\mathrm{g}(\mathcal{S}_1, x_1), \mathrm{g}(\mathcal{S}_2, x_2)) \mid x_1, x_2 \in \mathbb{R}^n : x_1 + x_2 = x \,\} \le 1$$

By positive homogeneity of the gauge functions, the equivalence between the inequalities implies eq. (61). $\square$

Proposition A.20 ([46, Theorem 3.7])**.** Consider the convex subsets $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{R}^n$, their *inverse sum*

$$\mathcal{S}_1 \,\sharp\, \mathcal{S}_2 = \bigcup_{\lambda \in \Delta^1} \lambda_1 \mathcal{S}_1 \cap \lambda_2 \mathcal{S}_2$$

is convex.

Proposition A.21. Consider the closed convex subsets $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{R}^n$, The following holds

$$(\mathcal{S}_1 + \mathcal{S}_2)^\circ = \mathcal{S}_1^\circ \,\sharp\, \mathcal{S}_2^\circ.$$

Proof. By proposition A.17, we have $\delta^*(y|\mathcal{S}_1 + \mathcal{S}_2) = \delta^*(y|\mathcal{S}_1) + \delta^*(y|\mathcal{S}_2)$. Therefore, $y \in (\mathcal{S}_1 + \mathcal{S}_2)^\circ$ if and only if there exists $0 < \lambda < 1$ such that $y \in \lambda \mathcal{S}_1^\circ$ and $y \in (1 - \lambda)\mathcal{S}_2^\circ$. $\square$

Proposition A.22. Given closed convex sets $\mathcal{S}_1, \mathcal{S}_2$ containing the origin, the following holds:

$$\mathrm{g}(\mathcal{S}_1 \,\sharp\, \mathcal{S}_2, x) = \mathrm{g}(\mathcal{S}_1, x) + \mathrm{g}(\mathcal{S}_2, x).$$

Proof. It follows from proposition A.21, proposition A.4 and proposition A.17. $\square$

Proposition A.23. Given nonempty closed convex sets $\mathcal{S}_1, \mathcal{S}_2$, the following holds:

$$\delta^*(y|\mathcal{S}_1 \,\sharp\, \mathcal{S}_2) = \left[\delta^*(\cdot|\mathcal{S}_1) \,\sharp\, \delta^*(\cdot|\mathcal{S}_2)\right](y).$$

Proof. It follows from proposition A.21, proposition A.4 and proposition A.19. $\square$

We have the following property for the Minkowski functions of linear preimages and support functions of linear images.

Proposition A.24. Given a matrix $A \in \mathbb{R}^{n_1 \times n_2}$ and a closed convex subset $\mathcal{S} \subseteq \mathbb{R}^{n_1}$ containing the origin, for all $x \in \mathbb{R}^{n_2}$, the following holds:

$$\mathrm{g}(A^{-1}\mathcal{S}, x) = \mathrm{g}(\mathcal{S}, Ax). \tag{62}$$

Proof. This is a direct consequence of proposition A.2 with $\alpha = 1$ and $f(x) = Ax$. □

Proposition A.25 ([45, Corollary 11.24(c)] or [46, Corollary 16.3.1])**.** Given a matrix $A \in \mathbb{R}^{n_1 \times n_2}$ and a nonempty closed convex set $\mathcal{S} \subseteq \mathbb{R}^{n_2}$, for all $y \in \mathbb{R}^{n_1}$, the following holds:

$$\delta^*(y|A\mathcal{S}) = \delta^*(A^\top y|\mathcal{S}). \tag{63}$$

Corollary A.2. Given a projection $\pi_{I,n}$ on the coordinates $I \subseteq [n]$ and a nonempty closed convex set $\mathcal{S} \subseteq \mathbb{R}^{n_2}$, the following holds:

$$\delta^*(y|\pi_{I,n}(\mathcal{S})) = \delta^*(\pi_{I,n}^\top(y)|\mathcal{S}).$$

The following properties show the relation between the linear image of a convex set and its polar.

Proposition A.26 ([46, Corollary 16.3.2])**.** For any convex set $\mathcal{S}$ and linear map A, the following equalities holds:

$$(A\mathcal{S})^\circ = A^{-\top}\mathcal{S}^\circ,$$
$$(A^{-1}(\operatorname{cl}\mathcal{S}))^\circ = \operatorname{cl}(A^\top(\mathcal{S}^\circ)).$$

Figure 9 illustrates the relation between four propositions, showing that either of these four is a consequence of the three other ones.

$$\begin{array}{ccc} \mathrm{g}(A^{-1}\mathcal{S}, x) & = & \mathrm{g}(\mathcal{S}, Ax) \\ \| & & \| \\ \delta^*(x|A^\top\mathcal{S}^\circ) & = & \delta^*(Ax|\mathcal{S}^\circ) \end{array}$$

Figure 9: Relation between proposition A.4, proposition A.24, proposition A.25 and proposition A.26.

Proposition A.26 plays a prominent role in Section 3.3. The following example provides a geometric intuition for it.

Example A.1. We can rewrite $(A\mathcal{S})^\circ = A^{-\top}(\mathcal{S}^\circ)$ as follows:

$$x \in (A\mathcal{S})^\circ \iff A^\top x \in \mathcal{S}^\circ \tag{64}$$

Let $\mathcal{S}$ be a sphere of radius 1 centered at $(0, 0, 2)$ and A be an orthogonal projection matrix which range equal to the plane $z = 0$. With these choices of $\mathcal{S}$ and A, the set $A\mathcal{S}$ is the circle of radius 1 of the plane $z = 0$ centered at the origin. This is illustrated by the Figure 10. Let x be the vector normal to the blue line shown in Figure 10. By (64), x is in the polar of the circle, if and only if $A^\top x$, which is the vector normal brown plane shown in Figure 10, is in the polar of the sphere.

A.4 Polyhedral computation

Hyperplanes and halfspaces are respectively denoted by

$$\overline{\mathcal{H}}_{a,\alpha} = \{\, x \mid a^\top x = \alpha \,\} \qquad\qquad \mathcal{H}_{a,\alpha} = \{\, x \mid a^\top x \le \alpha \,\}$$

The intersection of finitely many halfspaces is called a *polyhedral convex set* or a *polyhedron*. The convex hull of finitely many points is called a *polytope*.

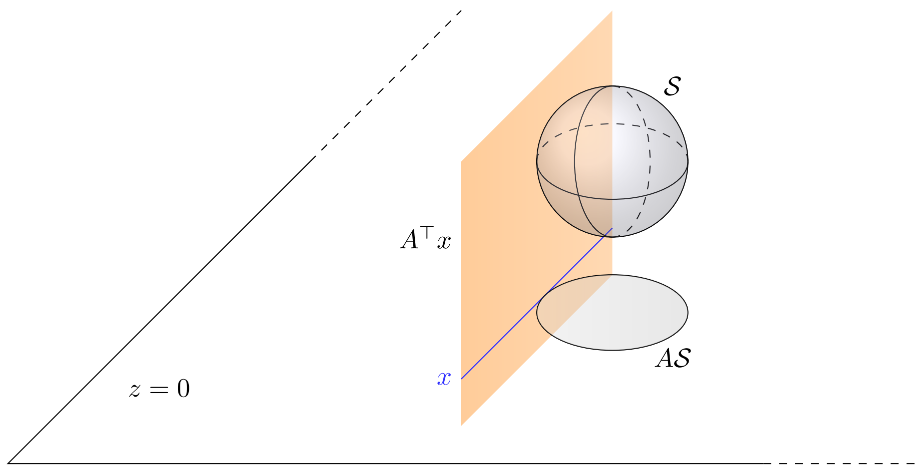


Figure 10: Illustration for Example A.1.

A.4.1 $\mathcal{H}$-representation and $\mathcal{V}$-representation

Theorem A.1 (Weyl-Minkowski)**.** For a subset $\mathcal{P}$ of $\mathbb{R}^n$, the following statements are equivalent

- $\mathcal{P}$ has an *$\mathcal{H}$-representation*, that is, there are s vectors $a_i \in \mathbb{R}^n$ and s scalars $\alpha_i \in \mathbb{R}$ such that

$$\mathcal{P} = \{ x \mid a_i^\top x \leq \alpha_i, \forall i \in [s] \} = \bigcap_{i \in [s]} \mathcal{H}_{a_i, \alpha_i}. \tag{65}$$

- $\mathcal{P}$ has a *$\mathcal{V}$-representation*, that is, there is finitely many vectors $v_1, \ldots, v_s \in \mathbb{R}^n$, $r_1, \ldots, r_t \in \mathbb{R}^n$ such that

$$\mathcal{P} = \mathrm{conv}\{v_1, \ldots, v_s\} + \mathrm{cone}\{r_1, \ldots, r_t\}. \tag{66}$$

Definition A.9. The set of *$\mathcal{H}$-representations* of a polyhedron $\mathcal{P} \subseteq \mathbb{R}^n$, denoted $\mathcal{H}_{\mathrm{rep}}(\mathcal{P})$ is the family of sets of pairs $(a_i, \alpha_i)_{i=1}^s \subseteq \mathbb{R}^{n+1}$ such that (65) holds.

The $\mathcal{H}$-representation of polyhedron gives the gauge function (see definition A.2).

Proposition A.27. Consider a polyhedron $\mathcal{P}$ with $\mathcal{H}$-representation given by $(a_i, \alpha_i)_{i=1}^s \in \mathcal{H}_{\mathrm{rep}}(\mathcal{P})$. If $\mathcal{P}$ contains the origin in its interior then

$$\mathrm{g}(\mathcal{P}, x) = \max_{i \in [s]} \langle a_i / \alpha_i, x \rangle.$$

Proof. As the origin is in the interior of $\mathcal{P}$, $\alpha_i > 0$ for $i = 1, \ldots, s$. Therefore, $\mathrm{g}(\mathcal{H}_{a_i, \alpha_i}, x) = \langle a_i / \alpha_i, x \rangle$. By theorem A.1, $\mathcal{P}$ is the intersection of the hyperplanes, see (65). The result follows from proposition A.12. $\square$

The following propositions are consequences of proposition A.12 and proposition A.24 respectively.

Proposition A.28. Consider polyhedra $\mathcal{P}_1, \mathcal{P}_2$, the concatenation of a $\mathcal{H}$-representation for each set is a $\mathcal{H}$-representation for their intersection $\mathcal{P}_1 \cap \mathcal{P}_2$.

Proposition A.29. Consider a polyhedron $\mathcal{P}$ with $\mathcal{H}$-representation given by $(a_i, \alpha_i)_{i=1}^s \in \mathcal{H}_{\mathrm{rep}}(\mathcal{P})$ and a matrix $A \in \mathbb{R}^{n \times m}$. A $\mathcal{H}$-representation of the preimage of $A^{-1}\mathcal{P}$ is given by

$$(A^\top a_i, \alpha_i)_{i=1}^s \in \mathcal{H}_{\mathrm{rep}}(A^{-1}\mathcal{P}).$$

Definition A.10. The set of *$\mathcal{V}$-representations* of a polyhedral cone $\mathcal{P} \subseteq \mathbb{R}^n$, denoted $\mathcal{V}_{\mathrm{rep}}(\mathcal{P})$ is the family of sets of vectors $(v_i)_{i=1}^s \subseteq \mathbb{R}^n$, $(r_i)_{i=1}^t) \subseteq \mathbb{R}^n$ such that (66) holds.

By proposition A.4 and proposition A.33, the following proposition follows from proposition A.27.

Proposition A.30. Consider a nonempty polyhedron $\mathcal{P}$ with $\mathcal{V}$-representation given by $v_1, \ldots, v_s \in \mathbb{R}^n$, $r_1, \ldots, r_t \in \mathbb{R}^n$. The support function of $\mathcal{P}$ is given by

$$\delta^*(x|\mathcal{P}) = \begin{cases} \infty & \text{if } \exists i \in [t] : \langle r_i, x\rangle > 0, \\ \max_{i\in[s]}\langle v_i, x\rangle & \text{otherwise.} \end{cases}$$

As a consequence of proposition A.30, the following propositions follow from proposition A.17 and proposition A.25 respectively.

Proposition A.31. Consider polyhedra $\mathcal{P}_1, \mathcal{P}_2$, with $\mathcal{V}$-representations $v_1^k, \ldots, v_{s_k}^k \in \mathbb{R}^n$, $r_1^k, \ldots, r_{t_k}^k \in \mathbb{R}^n$ for $\mathcal{P}_k$. The pairwise sum of the vertices $v_i^1 + v_j^2$ for $i \in [s_1], j \in [s_2]$ and concatenation of the vectors $(r_i^k)_{i=1}^{t_k}$ for $k = 1, 2$ form a $\mathcal{V}$-representation for the Minkowski sum $\mathcal{P}_1 + \mathcal{P}_2$.

Proof. It follows from proposition A.17, noticing that

$$\max_{i\in[s_1]}\langle v_i^1, x\rangle + \max_{i\in[s_2]}\langle v_i^2, x\rangle = \max_{i\in[s_1]}\max_{j=1}^{s_2}\langle v_i^1 + v_j^2, x\rangle.$$

□

Proposition A.32. Consider a polyhedron $\mathcal{P}$ with $\mathcal{V}$-representation given by $((v_i)_{i=1}^s, (r_i)_{i=1}^t) \in \mathcal{V}_{\text{rep}}(\mathcal{P})$ and a matrix $A \in \mathbb{R}^{m\times n}$. A $\mathcal{V}$-representation of the image $A\mathcal{P}$ is given by

$$((Av_i)_{i=1}^s, (Ar_i)_{i=1}^t) \in \mathcal{V}_{\text{rep}}(A\mathcal{P}).$$

Similarly to Corollary A.2, we have the following corollary.

Corollary A.3. Consider a polyhedron $\mathcal{P}$ with $\mathcal{V}$-representation given by $((v_i)_{i=1}^s, (r_i)_{i=1}^t) \in \mathcal{V}_{\text{rep}}(\mathcal{P})$ and a subset $I \subseteq [n]$. A $\mathcal{V}$-representation of the projection of $\mathcal{P}$ on I is given by

$$((\pi_{I,n}v_i)_{i=1}^s, (\pi_{I,n}r_i)_{i=1}^t) \in \mathcal{V}_{\text{rep}}(\pi_{I,n}(\mathcal{P})).$$

The $\mathcal{V}$-representation of a polytope containing the origin in its interior can directly be obtained from the $\mathcal{H}$-representation.

Proposition A.33. Consider a polytope $\mathcal{P}$ with $\mathcal{H}$-representation given by $(a_i, \alpha_i)_{i=1}^s \in \mathcal{H}_{\text{rep}}(\mathcal{P})$. If $\mathcal{P}$ contains the origin in its interior then $(a_i/\alpha_i)_{i=1}^s \in \mathcal{V}_{\text{rep}}(\mathcal{P}^\circ)$.

As a consequence, we can complete proposition A.7 which only gave the polar of the p-norm for $1 < p < \infty$. The following corollary shows that the polar of a cross-polytope is a hypercube.

Corollary A.4. The polar of the unit ball of the 1-norm is the unit ball of the ∞-norm.

A.4.2 Containment and inclusion

Given a point x, its containment in a polyhedron can be checked using either of the following two properties.

Proposition A.34. A point x belongs to a polyhedron $\mathcal{P}$ defined by (65) if and only if $\langle a_i, x\rangle \le \alpha_i$ for $i = 1, \ldots, s$.

Proposition A.35. A point x belongs to a polyhedron $\mathcal{P}$ defined by (66) if and only if there exists $(\lambda_1 \ldots, \lambda_s) \in \Delta^{s-1}$ and $\gamma_1, \ldots, \gamma_t \in \mathbb{R}_+$ such that

$$x = \sum_{i=1}^{s} \lambda_i v_i + \sum_{i=1}^{t} \gamma_i r_i. \tag{67}$$

The inclusion of a ray in a polyhedron can be checked using either of the following two properties.

Proposition A.36. There exists a translation of the ray $\{ \lambda r \mid \lambda \geq 0 \}$ that is included in a polyhedron $\mathcal{P}$ defined by (65) if and only if $\langle a_i, r \rangle \leq 0$ for $i = 1, \ldots, s$.

Proposition A.37. There exists a translation of the ray $\{ \lambda r \mid \lambda \geq 0 \}$ that is included in a polyhedron $\mathcal{P}$ defined by (66) if and only if there exists $\gamma_1, \ldots, \gamma_t \in \mathbb{R}_+$ such that

$$r = \sum_{i=1}^{t} \gamma_i r_i.$$

Given a halfspace $\mathcal{H}_{a,\alpha}$, the inclusion of a polyhedron in the halfspace can be checked using either of the following two properties.

Proposition A.38. A polyhedron $\mathcal{P}$ defined by (66) is included in a halfspace $\mathcal{H}_{a,\alpha}$ if and only if $\langle a, v_i \rangle \leq \alpha$ for $i = 1, \ldots, s$, $\langle a, r_i \rangle \leq 0$ for $i = 1, \ldots, t$ and either $s > 0$ or $t = 0$ or $\alpha \geq 0$.

Proposition A.39. A polyhedron $\mathcal{P}$ defined by (65) is included in a halfspace $\mathcal{H}_{a,\alpha}$ if and only if there exists $(\lambda_1 \ldots, \lambda_s) \in \mathbb{R}_+$ such that

$$\alpha \geq \sum_{i=1}^{s} \lambda_i \alpha_i$$

and

$$a = \sum_{i=1}^{s} \lambda_i a_i.$$

A.4.3 Representation conversion and projection

The complexity of algorithms is commonly described using only the input size (in bits) and not the output size. This is not appropriate for representation conversion and polyhedral projection algorithms that need to list the facets (resp. vertices, rays) of a polyhedron. Indeed, the output size itself might not be polynomial in terms of the input size. We say that an algorithm is *output-sensitive* if it is in P-TIME in both the input and output size.

The problem of computing the $\mathcal{H}$-representation (rep. $\mathcal{V}$-representation) given the $\mathcal{V}$-representation (resp. $\mathcal{H}$-representation) is the *representation conversion* problem. It was shown that none of the known algorithms for the representation conversion are output-sensitive [6] but it is still unknown whether there exists an output-sensitive algorithm for this problem.

An algorithm with a large output might produce its output incrementally without needing to store its output to continue. Therefore it might run in P-SPACE in its input size even if its output size is exponential in its input size. Such algorithm is said to be *compact* [18]. This is the case of the reversed search vertex enumeration algorithm implemented in LRS [5].

Given a $\mathcal{V}$-representation $((v_i)_{i=1}^s, (r_i)_{i=1}^t) \in \mathcal{V}_{\text{rep}}(\mathcal{P})$ of a polyhedron $\mathcal{P}$, a $\mathcal{H}$-representation of a lifted polyhedron in (x, λ, γ) is given by eq. (67). The $\mathcal{H}$-representation of the polyhedron can be obtained by eliminating λ and γ from this $\mathcal{H}$-representation. Eliminating variables from a polyhedron is called *polyhedral projection*. That is, we have just provided a polynomial reduction of representation conversion to polyhedral projection of $\mathcal{H}$-representation. A more sophisticated reduction, relying on the Farkas lemma, only requires computing the representation conversion problem for a polyhedra of dimension equal to the number of eliminated dimensions. This method is referred to as *block-elimination* and is implemented in CDD [19].

The family of polyhedra has the important property of being closed under projection:

Theorem A.2 (Fourier-Motzkin)**.** The projection of a polyhedral set on a linear subspace is a polyhedral set.

This is proved constructively by exhibiting the Fourier-Motzkin elimination which, given the $\mathcal{H}$-representation of a polyhedron, provides a procedure to compute the $\mathcal{H}$-representation of the polyhedron obtained by eliminating one coordinate. Even if there exists heuristics to limit this, the Fourier-Motzkin elimination procedure generally generates a redundant $\mathcal{H}$-representation. Removing these redundant halfspaces is crucial when eliminating several coordinates to avoid a prohibitive number of redundant halfspaces.

Polyhedral projection of a polyhedron in its $\mathcal{V}$-representation is simple. As shown by Corollary A.3, it boils down to dropping the eliminated coordinates of vertices and rays. We can see that polyhedral projection of an $\mathcal{H}$-representation can be achieved by two representation conversion and the polyhedral projection of $\mathcal{V}$-representation. This provides a polynomial reduction of polyhedral projection of $\mathcal{H}$-representation to representation conversion.

This shows that representation conversion is essentially "as hard" as polyhedral projection.

A.4.4 Zonotopes

We saw in proposition A.31 that taking the Minkowski sum of two polytopes increase the size of the $\mathcal{V}$-representation quadratically (even if some vertices may be redundant). One way to go around this issue is to define the following representation.

Definition A.11. A set of vectors $c, (v_i)_{i=1}^z$ is a *$\mathcal{Z}$-representation* of a polyhedron $\mathcal{P}$, denoted $(c, (v_i)_{i=1}^z) \in \mathcal{Z}_{\text{rep}}(\mathcal{P})$, if

$$\mathcal{P} = \{c\} + \sum_{i=1}^{z} \operatorname{conv}(-v_i, v_i).$$

Note that not all polytopes have a $\mathcal{Z}$-representation. The polytopes with a $\mathcal{Z}$-representation are called *Zonotopes*.

Proposition A.40. Consider a nonempty zonotope $\mathcal{P}$ with $\mathcal{Z}$-representation given by $(c, (v_i)_{i=1}^z) \in \mathcal{Z}_{\text{rep}}(\mathcal{P})$. The support function of $\mathcal{P}$ is given by

$$\delta^*(x|\mathcal{P}) = \langle c, x\rangle + \sum_{i=1}^{z} |\langle z_i, x\rangle|.$$

As a consequence, the following propositions follow from proposition A.17 and proposition A.25.

Proposition A.41. Consider zonotopes $\mathcal{P}_1, \mathcal{P}_2$, with $\mathcal{Z}$-representations $(c^k, (v_i^k)_{i=1}^z) \in \mathcal{Z}_{\text{rep}}(\mathcal{P}_k)$. The tuple made of $c^1 + c^2$ and the concatenation of v^1 and v^2 is a $\mathcal{Z}$-representation for the Minkowski sum $\mathcal{P}_1 + \mathcal{P}_2$.

Proposition A.42. Consider a zonotope $\mathcal{P}$ with $\mathcal{Z}$-representation given by $(c, (v_i)_{i=1}^z) \in \mathcal{Z}_{\text{rep}}(\mathcal{P})$ and a matrix $A \in \mathbb{R}^{m\times n}$. A $\mathcal{Z}$-representation of the image $A\mathcal{P}$ is given by

$$(Ac, (Av_i)_{i=1}^z) \in \mathcal{Z}_{\text{rep}}(A\mathcal{P}).$$

We see with these proposition that the zonotope family is closed under Minkowski sum and image (hence also projection). However, it is not closed under intersection or convex hull (otherwise any polytope would be a zonotope).

A.5 Ellipsoids

To obtain the polar of $\mathcal{E}_P$, we first develop the following property of the conjugate of quadratic forms.

Proposition A.43. Given a positive definite matrix $P \in \mathcal{S}_{++}^n$, the conjugate of the convex function $f(x) = x^\top Px/2$ is the convex function $f^*(y) = y^\top P^{-1}y/2$.

Proof. By corollary A.1, $y = Px$ if and only if $x = \nabla f^*(y)$. Therefore, $\nabla f^*(y) = P^{-1}y$ and $f^*(y) = y^\top P^{-1}y/2$. $\square$

We can deduce the following.

Proposition A.44. Given a positive definite matrix $P \in \mathcal{S}^n_{++}$, the polar of the ellipsoid $\mathcal{E}_P$ is $\mathcal{E}_P^\circ = \mathcal{E}_{P^{-1}}$.

Proof. Using (39), proposition A.6 and proposition A.43, we see that the polar of the gauge function is $\mathrm{g}(\mathcal{E}_P^\circ, x) = \mathrm{g}^\circ(\mathcal{E}_P, x) = \sqrt{x^\top P^{-1} x}$. □

We deduce the following from proposition A.44 and proposition A.4.

Proposition A.45. Given a positive definite matrix $P \in \mathcal{S}^n_{++}$, the support function of the ellipsoid $\mathcal{E}_P$ is $\delta^*(y|\mathcal{E}_P) = \sqrt{y^\top P^{-1} y}$.

A.5.1 Operations

As the family of ellipsoids is invariant under polarity operation, the six operations can be considered by pairs $(\cap, \mathrm{conv} \cup)$, $(+, \sharp)$ and (A, A^{-1}). The family of ellipsoids is either invariant under both operations of the pair or none.

The intersection and convex hull of the union of ellipsoid is not necessarily an ellipsoid, but it is always a *piecewise semi-ellipsoid*, that we study in Section 4.3.

The sum and inverse sum of ellipsoids is not ellipsoidal in general. By proposition A.17 and proposition A.45, we have that

$$\delta^*(y|\mathcal{E}_{P_1} + \mathcal{E}_{P_2}) = \sqrt{y^\top P_1 y} + \sqrt{y^\top P_2 y}$$

which is not a quadratic form in general. The following example provides a simple counterexample of the invariance of the family of ellipsoids under sum and inverse sum.

Example A.2. Consider the semi-ellipsoids $\mathcal{E}_{e_1 e_1^\top}$ and $\mathcal{E}_{e_2 e_2^\top}$. By proposition A.22 and (39), we have

$$\mathrm{g}(\mathcal{E}_{e_1 e_1^\top} \sharp \mathcal{E}_{e_2 e_2^\top}, x) = \sqrt{x^\top e_1 e_1^\top x} + \sqrt{x^\top e_2 e_2^\top x} = |x_1| + |x_2|.$$

Hence $\mathcal{E}_{e_1 e_1^\top} \sharp \mathcal{E}_{e_2 e_2^\top}$ is the unit ball of the 1-norm and, by Corollary A.4, its polar is the ∞-norm.

We have the following property for the linear preimage of an ellipsoid.

Proposition A.46. Given an ellipsoid $\mathcal{E}_P \subseteq \mathbb{R}^n$ and a matrix $A \in \mathbb{R}^{n \times m}$, the following holds:

$$A^{-1}\mathcal{E}_P = \mathcal{E}_{A^\top P A}.$$

Proof. This is a consequence of (39) and proposition A.24. □

In view of proposition A.45 and proposition A.25, the linear image of an ellipsoid is more naturally obtained from P^{-1} and its polar.

Proposition A.47. Given an ellipsoid $\mathcal{E}_P \subseteq \mathbb{R}^n$ and a matrix $A \in \mathbb{R}^{m \times n}$, the following holds:

$$\delta^*(y|A\mathcal{E}_P) = y^\top A P^{-1} A^\top y \tag{68}$$
$$A\mathcal{E}_P^\circ = \mathcal{E}_{AP^{-1}A^\top}. \tag{69}$$

Moreover, if $AP^{-1}A^\top$ is invertible, we have

$$A\mathcal{E}_P = \mathcal{E}_{(AP^{-1}A^\top)^{-1}}. \tag{70}$$

Proof. The equation (68) follows from proposition A.45 and proposition A.25. By proposition A.4 and proposition A.45, equation (69) is obtained from (68). The set $\mathcal{E}_{AP^{-1}A^\top}$ is a semi-ellipsoid but it is an ellipsoid only if $AP^{-1}A^\top$ is invertible. If it is an ellipsoid, (70) follows from proposition A.44 and (69). □

A.5.2 Containment and inclusion

As a consequence of proposition A.9 and (39), the inclusion between ellipsoids is verified using the following property.

Proposition A.48. Consider two ellipsoids $\mathcal{E}_{P_1}, \mathcal{E}_{P_2}$. The inclusion $\mathcal{E}_{P_1} \subseteq \mathcal{E}_{P_2}$ is equivalent to $P_1 \succeq P_2$.

Proposition A.49 (Schur complement)**.** Consider the matrices $A \in \mathbb{R}^{n_1 \times n_1}, B \in \mathbb{R}^{n_1 \times n_2}, C \in \mathbb{R}^{n_2 \times n_1}, D \in \mathbb{R}^{n_2 \times n_2}$. Let

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

If D is invertible, the *Schur complement* of the block D of the matrix M is defined by

$$M/D = A - BD^{-1}C.$$

If A is invertible, the *Schur complement* of the block A of the matrix M is defined by

$$M/A = D - BA^{-1}C.$$

We have the following properties of the Schur complement: If D is invertible then the matrix M is positive definite if and only if D and M/D are positive definite. If A is invertible then the matrix M is positive definite if and only if A and M/A are positive definite. Moreover, if $A \in \mathcal{S}^{n_1}, D \in \mathcal{S}^{n_2}$ and $C = B^\top$, we have the following properties. If D is positive definite then the matrix M is positive semidefinite if and only if M/D is positive semidefinite. If A is positive definite then the matrix M is positive semidefinite if and only if M/A is positive semidefinite.

By definition of the gauge function and proposition A.49, we have the following two conditions for the containment of a point in an ellipsoid.

Proposition A.50. Given a point $x \in \mathbb{R}^n$ and an ellipsoid $\mathcal{E}_P$, the membership $x \in \mathcal{E}_P$ is equivalent to any of the following three equivalent conditions

$$x^\top P x \leq 1, \tag{71}$$

$$\begin{bmatrix} 1 & x^\top \\ x & P^{-1} \end{bmatrix} \succeq 0, \tag{72}$$

$$P^{-1} \succeq xx^\top. \tag{73}$$

Proof. The condition (71) is by definition of the ellipsoid, see (38). The condition (72) follows from (71) and proposition A.49. The condition (73) follows from (72) and proposition A.49 but we provide here an alternative proof that gives a more geometric interpretation. As $\mathcal{E}_P$ is a symmetric convex set, $x \in \mathcal{E}_P$ if and only if $\operatorname{conv}(\{-x, x\}) \subseteq \mathcal{E}_P$. By proposition A.30, we have $\delta^*(y|\operatorname{conv}(\{-x, x\})) = |\langle x, y\rangle| = \sqrt{y^\top x x^\top y}$. By proposition A.45, $\delta^*(y|\mathcal{E}_P) = \sqrt{x^\top P^{-1} x}$, Therefore, by proposition A.10, we have (73). $\square$

By proposition A.44 and proposition A.49, we have the following two conditions for the inclusion of an ellipsoid in a halfspace.

Proposition A.51. Given a halfspace $\mathcal{H}_{a,\alpha} \subseteq \mathbb{R}^n$ with $\alpha > 0$ and an ellipsoid $\mathcal{E}_P$, the inclusion $\mathcal{E}_P \subseteq \mathcal{H}_{a,\alpha}$ is equivalent to the following three equivalent conditions

$$a^\top P^{-1} a \leq \alpha^2, \tag{74}$$

$$\begin{bmatrix} \alpha^2 & a^\top \\ a & P \end{bmatrix} \succeq 0, \tag{75}$$

$$P \succeq \alpha^{-2} a a^\top. \tag{76}$$

Proof. The condition (74) follows proposition A.33 and proposition A.44. The condition (75) follows from (74) and proposition A.49. The condition (76) follows from (75) and proposition A.49 but we provide here an alternative proof that gives a more geometric interpretation. As $\mathcal{E}_P$ is a symmetric convex set, $\mathcal{E}_P \subseteq \mathcal{H}_{a,\alpha}$ if and only if $\mathcal{E}_P \subseteq \mathcal{H}_{a,\alpha} \cap \mathcal{H}_{-a,\alpha}$. By proposition A.27, we have $\mathrm{g}(\mathcal{H}_{a,\alpha} \cap \mathcal{H}_{-a,\alpha}, x) = |\langle x, a/\alpha \rangle| = \sqrt{x^\top a a^\top x}/\alpha$. By proposition A.45, $\delta^*(y|\mathcal{E}_P) = \sqrt{x^\top P^{-1} x}$, Therefore, by (39) and proposition A.8, we have (76). $\square$

A.5.3 Volume and semi-axis

An ellipsoid $\mathcal{E}_P$ can be represented as the rotation of an ellipsoid of the form $\mathcal{E}_{e_1 e_1^\top / a_1^2 + \ldots + e_n e_n^\top / a_n^2}$. Geometrically, the numbers a_i represent the length of the semi-axis of the ellipsoid. The length of the semi-axes is an invariant of the ellipsoids under rotation.

The volume of an ellipsoid does not depend on the rotation and can be expressed in terms of this invariant.

Proposition A.52. The volume of an ellipsoid with semi-axis of length $a_1, \ldots, a_n$ is given by

$$\frac{\pi^{n/2}}{\Gamma(n/2+1)} \prod_{i=1}^{n} a_i.$$

The numbers a_i are the inverse of square root of the eigenvalues of the matrix P, in other words, they are the square root of the eigenvalues of the inverse of P. The following two properties follow from this observation.

Proposition A.53. The volume of an ellipsoid $\mathcal{E}_P$ is given by

$$\frac{\pi^{n/2}}{\Gamma(n/2+1)\sqrt{\det(P)}}.$$

Proposition A.54. The sum of the squared length of the semi-axes of an ellipsoid $\mathcal{E}_P$ is given by $\operatorname{Tr} P^{-1}$.