LOW-COMPLEXITY METHOD FOR HYBRID MPC WITH LOCAL GUARANTEES[*]

DAMIAN FRICK[†], ANGELOS GEORGHIOU[‡], JUAN L. JEREZ[†§], ALEXANDER DOMAHIDI[‡¶], AND MANFRED MORARI[†]

Abstract. Model predictive control problems for constrained hybrid systems are usually cast as mixed-integer optimization problems (MIP). However, commercial MIP solvers are designed to run on desktop computing platforms and are not suited for embedded applications which are typically restricted by limited computational power and memory. To alleviate these restrictions, we develop a novel low-complexity, iterative method for a class of non-convex, non-smooth optimization problems. This class of problems encompasses hybrid model predictive control problems where the dynamics are piece-wise affine (PWA). We give conditions such that the proposed algorithm has fixed points and show that, under practical assumptions, our method is guaranteed to converge locally to local minima. This is in contrast to other low-complexity methods in the literature, such as the non-convex alternating directions method of multipliers (ADMM), for which no such guarantees are known for this class of problems. By interpreting the PWA dynamics as a union of polyhedra we can exploit the problem structure and develop an algorithm based on operator splitting procedures. Our algorithm departs from the traditional MIP formulation, and leads to a simple, embeddable method that only requires matrix-vector multiplications and small-scale projections onto polyhedra. We illustrate the efficacy of the method on two numerical examples, achieving good closed-loop performance with computational times several orders of magnitude smaller compared to state-of-the-art MIP solvers. Moreover, it is competitive with ADMM in terms of suboptimality and computation time, but additionally provides local optimality and local convergence guarantees.

1. Introduction. Many practical applications for control fall in the domain of hybrid systems, including applications such as active suspension control or energy management in automotives [2, 9] and applications in power electronics [13]. These applications require fast sampling times in the sub-second range, and the control algorithms need to be implemented on industrial, resource-constrained platforms, such as microcontrollers or re-configurable hardware. These systems are characterized by complex interactions between discrete and continuous behaviors that make them extremely challenging to control. In the last decades model predictive control (MPC) [26] has received widespread attention both in research and industry. It provides a systematic approach for controlling constrained hybrid systems, promising high control performance with minimal tuning effort.

For hybrid systems, mature modeling tools such as the mixed logical dynamical framework, see [5], are available. It defines rules for posing hybrid MPC as mixed-integer optimization problems (MIP). These MIPs can then be tackled with powerful commercial solvers such as CPLEX [19]. These solvers have focused on the solution of generic, large-scale, mixed-integer linear/quadratic optimization problems on powerful computing platforms, and thus have several drawbacks for embedded applications: (i) their code size is in the order of tens of megabytes; (ii) the algorithms require substantial working memory; (iii) they depend on numerical libraries that cannot be ported to most embedded platforms. This has restricted the applicability of hybrid

MPC to systems that can run on desktop computing platforms, with sampling times in the order of minutes or even hours.

Recently there have been efforts to enable hybrid MPC applications with faster dynamics to run on platforms with more limited computational power. Explicit MPC [1] is the method of choice for very small problems but quickly becomes intractable for increasing problem dimension. Therefore, methods that further exploit (i) the sparsity-inducing multistage structure of MPC problems, and (ii) the fact that MPC problems are parametric and solved in a receding horizon fashion, have been developed. Methods based on MIP reformulations have been investigated, see [3, 12, 20, 31], and have led to reduced computation times compared to general-purpose solvers. However, these methods are not fast enough for many practical applications.

Methods based on operator splitting allow to trade off computational complexity with optimality, i.e., they typically produce "good" solutions in a fraction of the time required even by tailored MIP solvers to find global optima. Furthermore, they allow the problem to be split into several small subproblems, making these methods especially suited for embedded applications and implementations on computing systems with limited resources. The alternating directions method of multipliers (ADMM) [7] is an example of such operator splitting methods. In [32], ADMM was used on various hybrid MPC examples achieving substantial gains in computational performance compared to general-purpose MIP solvers. Convergence guarantees are available for some special cases of non-convex ADMM [22, 25], however, neither optimality or convergence guarantees are available for the class of problems considered in this work. Other low-complexity methods based on operator splitting suited for non-convex MPC are [17, 18]. They are attractive because they offer computational performance similar to ADMM, but unlike ADMM they still retain some guarantees on optimality and convergence. Nonetheless, they do not address the class of problems considered here.

1.1. Contribution. In this paper, we develop an algorithm based on operator splitting to address hybrid MPC problems, that provides both the desired theoretical guarantees and computational speed. Our method exploits the multistage structure of the optimization problem, allowing to decompose the complex piece-wise affine (PWA) equality constraints into decoupled non-convex polyhedral constraints. By leveraging their polyhedral nature, we are able to derive guarantees on optimality and convergence. The main contributions of this paper are the following:

• We develop an iterative, numerical method to find local minima of a class of non-convex, non-smooth optimization problems that arise in MPC problems for systems with PWA dynamics. Our method is of low complexity and only requires matrix-vector multiplications and small-scale Euclidean projections onto convex polyhedra.

• We provide (i) conditions for the existence of fixed points of the proposed algorithm; (ii) show that the method converges locally to such fixed points; and (iii) prove that fixed points correspond to local minima. To the author's knowledge, for the class of problems considered, this is the first method of such simplicity that allows to give local optimality and convergence guarantees.

• The proposed algorithm is implemented using automatic code generation, directly targeting the embedded applications. This code generation tool is made publicly available via github [11]. We examine the method's performance on two numerical example and demonstrate computational speed-ups of several orders of magnitude, compared to conventional MIP solvers, with only a minor loss of optimality. We further show, that the method is competitive with other local methods such as ADMM.

1.2. Outline. In Section 2, we state the problem. In Section 3, we construct an operator whose fixed points correspond to local minima. In Section 4, we propose an algorithm based on this operator and discuss local optimality and local convergence of the method. In Section 5, we present numerical results. Proofs of auxiliary results are given in the appendix.

1.3. Notation & elementary definitions. For vectors $x, y \in \mathbb{R}^n$, $\|x\| := \sqrt{x^\top x}$ is the 2-norm and $\langle x, y \rangle := x^\top y$ the scalar product. For a closed set $\mathcal{C}$, $\mathcal{P}_\mathcal{C}(x) := \arg\min_{z \in \mathcal{C}} \|x - z\|$ is the Euclidean projection of x onto $\mathcal{C}$. The closed ϵ-ball around x is $\mathcal{B}_\epsilon(x) := \{z \mid \|x - z\| \leq \epsilon\}$, we use the shorthand $\mathcal{B}_\epsilon := \mathcal{B}_\epsilon(0)$. The *relative interior* is $\text{relint}\,\mathcal{C}$. By $\mathcal{C} + \{x\}$ we denote the Minkowski sum of $\mathcal{C}$ and the singleton x. The *characteristic function* of set $\mathcal{C}$ is $\chi_\mathcal{C}(x) := \{0 \text{ if } x \in \mathcal{C}\,; +\infty \text{ otherwise}\}$. An *operator* $T : \mathcal{X} \rightrightarrows \mathcal{Y}$ maps every point $x \in \mathcal{X}$ to a set $T(x) \subseteq \mathcal{Y}$. T is called *single-valued* if for all $x \in \mathcal{X}$: $T(x)$ is a singleton. A *zero* of T is a point $x \in \mathcal{X}$ such that $0 \in T(x)$, denoted by $x \in \text{zero}\,T$, where $\text{zero}\,T := \{x \mid 0 \in T(x)\}$. A *fixed point* of T is a point $x \in \mathcal{X}$ such that $x \in T(x)$, denoted by $x \in \text{fix}\,T$, where $\text{fix}\,T := \{x \mid x \in T(x)\}$.

2. Problem statement. In this work, we consider optimization problems that are *parametric* in $\theta \in \mathbb{R}^p$ and given as follows:

$$\tag{1a} p^\star(\theta) := \min_{z \in \mathcal{E} \cap \mathcal{Z}(\theta)} \tfrac{1}{2} z^\top H z + h^\top z \,,$$

with $H \in \mathbb{R}^{n \times n}$ positive definite, and $z := (z_1, \ldots, z_N) \in \mathbb{R}^n$, with $z_k \in \mathbb{R}^{n_k}$. The parametric nature implies that only parts of the problem data will change every time the solved is invoked. The non-convex set $\mathcal{Z}(\theta) \subseteq \mathbb{R}^n$ has the form:

$$\tag{1b} \mathcal{Z}(\theta) := \underset{k=1}{\overset{N}{\times}} \left(\bigcup_{i=1}^{m_k} \mathcal{Z}_{k,i}(\theta) \right),$$

where the sets $\mathcal{Z}_{k,i}(\theta) \subseteq \mathbb{R}^{n_k}$ are closed convex polyhedra that may depend linearly on θ, in the following way: $\mathcal{Z}_{k,i}(\theta) := \{z_k \in \mathbb{R}^{n_k} \mid G_{k,i} z_k = g_{k,i}(\theta)\,, F_{k,i} z_k \leq f_{k,i}(\theta)\}$, with matrices and vectors of appropriate dimensions, and $g_{k,i}(\theta)$, $f_{k,i}(\theta)$ affine functions of the parameter θ. For a given parameter θ, the Euclidean projection onto $\mathcal{Z}(\theta)$ can be evaluated very efficiently by decoupling it into simple convex projections, see Algorithm 2 in Section 4. This important feature of $\mathcal{Z}$ will be a key ingredient of the algorithm presented in Section 4. For simplicity of notation we will omit the parameter dependence of $\mathcal{Z}(\theta)$ in the remainder of this paper. Finally, the set $\mathcal{E}$ is an affine equality constrained set $\mathcal{E} := \{z \in \mathbb{R}^n \mid Az = b\} = \{Vv + \bar{v} \mid v \in \mathbb{R}^{n-m}\}$, with $A \in \mathbb{R}^{m \times n}$ full row rank, and an appropriate matrix $V \in \mathbb{R}^{n \times n - m}$ and vector $\bar{v} \in \mathbb{R}^n$ [27, p. 430, (15.14)], where V has full column rank by construction. We furthermore assume that $\mathcal{E} \cap \mathcal{Z}$ is non-empty, hence the problem is feasible.

2.1. Consensus form. Problem (1) can be re-written in the so-called consensus form by introducing copies y of z as follows

$$\tag{2} p^\star := \min_{z,y\,:\,z=y} \tfrac{1}{2} z^\top H z + h^\top z + \chi_\mathcal{E}(z) + \chi_\mathcal{Z}(y) \,,$$

which is a non-convex, non-smooth optimization problem. The constraint sets $\mathcal{E}$ and $\mathcal{Z}$ have been moved into the cost function using their characteristic functions $\chi_\mathcal{E}(z)$ and $\chi_\mathcal{Z}(y)$. This problem formulation ensures that the two sets $\mathcal{E}$ and $\mathcal{Z}$ can be decoupled. The consensus constraint $z = y$ encodes the coupling between z and y.

2.2. Hybrid Model Predictive Control. We are particularly interested in MPC problems where the discrete-time dynamics are given by a PWA function of the states and inputs. Such parametric hybrid MPC problems can be written as:

$$(3a) \qquad \min_{x_k, u_k} \quad \sum_{k=1}^{N} q_{k+1}(x_{k+1}) + r_k(u_k)$$

$$(3b) \qquad \text{s. t.} \quad x_{k+1} = A_{k,i} x_k + B_{k,i} u_k + c_{k,i} \quad \text{if } (x_k, u_k) \in \mathcal{C}_{k,i}, \ k = 1, \ldots, N,$$

$$(3c) \qquad \qquad x_1 = \theta,$$

for $i = 1, \ldots, m_k$ with m_k representing the number of regions of the PWA dynamics. Vector $x_k \in \mathbb{R}^{n_x}$ denotes the state of the system at time k, $u_k \in \mathbb{R}^{n_u}$ denotes the inputs applied to the system between times k and $k + 1$, and $\theta \in \mathbb{R}^{n_x}$ is the (parametric) initial state of the system. The objective terms q_{k+1} and r_k are strictly convex, quadratic functions. The sets $\mathcal{C}_{k,i}$ are non-empty closed convex polyhedra that define the partition of the PWA dynamics, and include state and input constraints.

Problems of the form (3) can be solved by introducing additional continuous and binary variables, and formulating an MIP using, e.g., the big-M reformulation [5] or more advanced formulations [33]. The resulting problems can then be solved using off-the-shelf commercial MIP solvers. Instead of formulating a MIP, in this work we transform Problem (3) into the form (1). In this way only $N n_x$ auxiliary continuous variables and *no binary variables* need to be introduced. For each time instant k we introduce a copy w_k of x_{k+1}, as follows: $w_k := A_{k,i} x_k + B_{k,i} u_k + c_{k,i}$, if $(x_k, u_k) \in \mathcal{C}_{k,i}$. This allows us to define the closed convex polyhedral sets

$$\mathcal{Z}_{1,\hat{i}}(\theta) := \left\{ (u_1, w_1) \mid w_1 = A_{1,\hat{i}} \theta + B_{1,\hat{i}} u_1 + c_{1,\hat{i}} \text{ and } (\theta, u_1) \in \mathcal{C}_{1,\hat{i}} \right\},$$
$$\mathcal{Z}_{k,i} := \left\{ (x_k, u_k, w_k) \mid w_k = A_{k,i} x_k + B_{k,i} u_k + c_{k,i} \text{ and } (x_k, u_k) \in \mathcal{C}_{k,i} \right\},$$

for $\hat{i} = 1, \ldots, m_1$ and for $i = 1, \ldots, m_k$, $k = 2, \ldots, N$. The non-convex constraint set $\mathcal{Z}$ now has the form of (1b) and is *decoupled* in the horizon N. For an equivalent formulation, we impose $x_{k+1} = w_k$ by defining a set of *coupling* equality constraints

$$\mathcal{E} := \bigtimes_{k=1}^{N} \left(\mathbb{R}^{n_u} \times \left\{ (w_k, x_{k+1}) \in \mathbb{R}^{2n_x} \mid x_{k+1} = w_k \right\} \right).$$

Remark 2.1. To ensure the strict convexity of Problem (1) equivalent to Problem (3), we impose the objective $\alpha_k q_{k+1}(x_{k+1}) + (1 - \alpha_k) q_{k+1}(w_k) + r_k(u_k)$ for each k and any $\alpha_k \in (0, 1)$. With the coupling constraint $w_k = x_{k+1}$ this ensures that the cost remains unchanged. Without loss of generality, we have used $\alpha_k = 0.5$ for all k.

3. KKT conditions for Problem (2). We investigate Karush-Kuhn-Tucker (KKT) conditions for Problem (2) and construct an operator K_ξ for finding particular KKT points. The properties and structure of this operator allow us to establish the computational efficiency of the method and show local optimality and convergence.

3.1. Generalized, regular & proximal KKT points. In the context of convex optimization the KKT conditions give rise to necessary and sufficient conditions for global optimality, under mild assumptions. In the non-convex, non-smooth case, however, they are in general neither necessary nor sufficient, even for local optimality. We use a notion of KKT points for instances of Problem (2) that provide necessary conditions for local optimality. Computing such KKT points is in general intractable. Therefore, we propose two inner approximations that are more computationally attractive. We denote KKT points by †, local optima by $^\circ$ and global optima by *.

A point $(z^\dagger, y^\dagger, \lambda^\dagger)$ is called a generalized KKT point [30, Theorem 3.34, p. 127] of Problem (2) if it satisfies

$$\text{(4a)} \qquad\qquad 0 \in \{Hz^\dagger + h\} + \mathcal{N}_{\mathcal{E}}(z^\dagger) - \{\lambda^\dagger\},$$

$$\text{(4b)} \qquad\qquad 0 \in \mathcal{N}_{\mathcal{Z}}(y^\dagger) + \{\lambda^\dagger\},$$

$$\text{(4c)} \qquad\qquad 0 = z^\dagger - y^\dagger,$$

where the regular normal cone $\hat{\mathcal{N}}_{\mathcal{C}}(z)$ and the (limiting) normal cone $\mathcal{N}_{\mathcal{C}}(z)$ of a set $\mathcal{C}$ are defined as in [29, Proposition 6.5, p. 200]:

$$\hat{\mathcal{N}}_{\mathcal{C}}(z) := \left\{ x \in \mathbb{R}^n \mid \limsup_{u \xrightarrow{c} z} \frac{\langle x, u - z \rangle}{\|u - z\|} \right\}, \text{ and } \mathcal{N}_{\mathcal{C}}(z) := \limsup_{x \xrightarrow{c} z} \hat{\mathcal{N}}_{\mathcal{C}}(x) = \partial \chi_{\mathcal{C}}(z).$$

In particular, the KKT conditions (4) imply that $z^\dagger = y^\dagger \in \mathcal{E} \cap \mathcal{Z}$. Since any KKT point $(z^\dagger, y^\dagger, \lambda^\dagger)$ satisfies primal feasibility, $z^\dagger = y^\dagger$, we will use the shorthand notation $(z^\dagger, \lambda^\dagger)$ for KKT points in the remainder of this paper. The KKT conditions (4) give rise to necessary conditions for local optimality as detailed below.

LEMMA 3.1 (Necessary optimality condition). *For every local minimum z° of Problem (1) there exist Lagrange multipliers $\lambda^\circ \in \mathbb{R}^n$ such that (z°, λ°) is a generalized KKT point of Problem (2).*

These KKT conditions are difficult to exploit due to the dependence on $\mathcal{N}_{\mathcal{Z}}(\cdot)$, which cannot be evaluated easily [15]. However, we can define two inner approximations that admit practical characterizations, resulting in what we call *regular* and *ξ-proximal* KKT points. These special cases of KKT points can be evaluated more efficiently and have important regularity properties. Furthermore, in Section 4.2, we will show that they provide sufficient (but no longer necessary) conditions for local optimality. We call a point $(z^\dagger, \lambda^\dagger)$ a *regular* KKT point of Problem (2), if it satisfies

$$\text{(5)} \qquad 0 \in \{Hz^\dagger + h\} + \hat{\mathcal{N}}_{\mathcal{E}}(z^\dagger) - \{\lambda^\dagger\} \text{ and } 0 \in \hat{\mathcal{N}}_{\mathcal{Z}}(z^\dagger) + \{\lambda^\dagger\}.$$

Note that we have replaced the normal cone $\mathcal{N}_{\mathcal{Z}}(z^\dagger)$ from (4b) with the regular normal cone $\hat{\mathcal{N}}_{\mathcal{Z}}(z^\dagger)$. Since $\hat{\mathcal{N}}_{\mathcal{Z}}(z^\dagger) \subseteq \mathcal{N}_{\mathcal{Z}}(z^\dagger)$, see [29, Proposition 6.5, p. 200], the set of regular KKT points is an inner approximation of the set of generalized KKT points. Given $\xi > 0$, we call $(z^\dagger, \lambda^\dagger)$ a *ξ-proximal* KKT point of Problem (2), if it satisfies

$$\text{(6a)} \qquad\qquad 0 \in \{Hz^\dagger + h\} + \hat{\mathcal{N}}_{\mathcal{E}}(z^\dagger) - \{\lambda^\dagger\}, \text{ and}$$

$$\text{(6b)} \qquad\qquad z^\dagger \in \mathcal{P}_{\mathcal{Z}}(z^\dagger - \tfrac{1}{\xi}\lambda^\dagger).$$

Compared to the definition of regular KKT points we have replaced the regular normal cone $\mathcal{N}_{\mathcal{Z}}(z^\dagger)$ with a condition on the projection, and we have introduced a positive scaling ξ using $z^\dagger \in \mathcal{P}_{\mathcal{Z}}(z^\dagger - \tfrac{1}{\xi}\lambda^\dagger) \Rightarrow -\tfrac{1}{\xi}\lambda^\dagger \in \hat{\mathcal{N}}_{\mathcal{Z}}(z^\dagger)$, from [29, Example 6.16, p. 212]. This directly implies that every ξ-proximal KKT point $(z^\dagger, \lambda^\dagger)$ is also a regular KKT point and therefore (6) is an inner approximation of (5).

Evaluating whether a pair (z, λ) is a regular KKT point has an exponential worst-case complexity, because representing $\hat{\mathcal{N}}_{\mathcal{Z}}(z)$ may be combinatorial and can quickly become intractable. In contrast, evaluating whether a pair (z, λ) is a ξ-proximal KKT point amounts to a projection that can be performed in polynomial-time. Furthermore, for the primal variables z, this inner approximation (6) of (5) can be made tight, as shown in the following proposition. Figure 1 illustrates the conceptual relationship between the different KKT points.
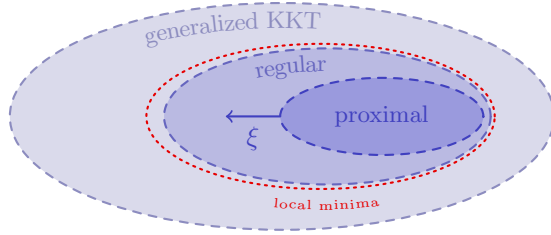
Fig. 1: The sets illustrate the relationship between different KKT points and represent points z for which there exist multipliers λ such that (z, λ) is a generalized, regular or proximal KKT point (dashed), as well as the set of local minima (dotted).

PROPOSITION 3.2. *Consider Problem (2).*
    i) For any regular KKT point $(z^\dagger, \lambda^\dagger)$, there exists a $\bar{\xi} > 0$ such that for all $\xi \geq \bar{\xi}$, $(z^\dagger, \lambda^\dagger)$ is a ξ-proximal KKT point.
    ii) Given any $\xi > 0$ and ξ-proximal KKT point $(z^\dagger, \lambda^\dagger)$, then $(z^\dagger, \lambda^\dagger)$ is also a regular KKT point.

Problem (2) has only finitely many local minima due to positive definiteness of H and the polyhedrality of $\mathcal{E} \cap \mathcal{Z}$. Therefore, given a $\xi > 0$ large enough, for every regular KKT point $(z^\dagger, \lambda^\dagger)$ there exists a $\mu^\dagger$ such that $(z^\dagger, \mu^\dagger)$ is a ξ-proximal KKT point.

3.2. Operator for proximal KKT points. The fact that set-membership evaluation of (6) is efficient, motivates the construction of an operator K_ξ that has the property that (i) it is cheap to evaluate and (ii) has zeros corresponding to ξ-proximal KKT points of Problem (2). We define a set-valued operator $K_\xi : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$:

$$(7a) \qquad\qquad K_\xi(s) := \{M_\xi s + c_\xi\} - \mathcal{P}_\mathcal{Z}(s),$$

with $\xi > 0$ the *proximal scaling*, and

$$(7b) \qquad M_\xi := \xi (\xi R - I)^{-1} R, \quad c_\xi := (\xi R - I)^{-1} (R(h + H\bar{v}) - \bar{v}),$$
$$(7c) \qquad R := V(V^\top H V)^{-1} V^\top,$$

where V and $\bar{v}$ as in Section 2. We will show that finding an s such that $0 \in K_\xi(s)$ is equivalent to finding a ξ-proximal KKT point $(z^\dagger, \lambda^\dagger)$. For K_ξ to be defined properly, $\xi R - I$ needs to be invertible. Since $R \succeq 0$, this holds for all

$$(8a) \qquad\qquad\qquad \xi > \left(\lambda_{\min}^+(R)\right)^{-1}.$$
$$(8b) \qquad\qquad \text{This implies } M_\xi \succeq 0 \text{ and } \lambda_{\min}^+(M_\xi) > 1,$$

where $\lambda_{\min}^+(R)$ is the smallest non-zero eigenvalue of R. For Problem (3), it can be shown that $\xi > \lambda_{\max}(H)$ is necessary and sufficient for satisfying both (8a) and (8b). Using (7a) we can show that a point $s^\dagger \in \text{zero } K_\xi$, called a zero of K_ξ, corresponds to a ξ-proximal KKT point $(z^\dagger, \lambda^\dagger)$ with $z^\dagger$, $\lambda^\dagger$ as follows:

$$(9) \qquad\qquad z^\dagger := M_\xi s^\dagger + c_\xi \text{ and } \lambda^\dagger := \xi(z^\dagger - s^\dagger),$$

where the precise relationship is given in Lemma 3.3.

LEMMA 3.3 (Zeros of operator K_ξ). *Given any $\xi > 0$. A point $s^\dagger \in \mathbb{R}^n$ satisfies $s^\dagger \in \operatorname{zero} K_\xi$ if and only if the pair $(z^\dagger, \lambda^\dagger)$ constructed through* (9) *is a ξ-proximal KKT point of Problem* (2).

Multiple points $s^\dagger \in \operatorname{zero} K_\xi$ can result in the same $z^\dagger$ (but different $\lambda^\dagger$), since M_ξ is only invertible when $\mathcal{E} = \mathbb{R}^n$. This means in particular that there can be multiple zeros of K_ξ corresponding to the same local optimum.

4. Solution method and theoretical guarantees. We present a method for finding ξ-proximal KKT points of Problem (2) by finding zeros of the operator K_ξ. The basis of this method is a fixed-point iteration, where every iteration is computationally inexpensive. We will show that every zero of K_ξ corresponds to a local minimum of Problem (2). Furthermore, under practical assumptions, the presented algorithm converges locally to such minima, almost always. By 'almost always' we mean that, for any local minimum that has corresponding fixed points, there exist at most a measure zero subset of fixed points from which the method may not converge.

4.1. Fixed-point algorithm. We define an operator $T_\xi : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, that has the zeros of K_ξ as its fixed points:

$$(10a) \qquad T_\xi(s) := s - W K_\xi(s) = s - W\left(M_\xi s + c_\xi - \mathcal{P}_\mathcal{Z}(s)\right),$$

with W invertible to ensure $\operatorname{fix} T_\xi = \operatorname{zero} K_\xi$. W is defined as

$$(10b) \qquad W := Q \begin{bmatrix} \frac{1}{2}\Lambda^{-1} & 0 \\ 0 & -I \end{bmatrix} Q^\top,$$

where $Q \in \mathbb{R}^{n \times n}$ is orthonormal and $\Lambda \in \mathbb{R}^{n-m \times n-m}$ diagonal, positive definite, such that $M_\xi = Q \operatorname{diag}(\Lambda, 0) Q^\top$ is the eigendecomposition of M_ξ. In addition, the structure of W will be used to prove local non-expansiveness of the operator T_ξ, see Lemma 4.5 in Section 4.4. The local non-expansiveness of T_ξ allows us to use the *Krasnoselskij iteration* [6, Theorem 3.2, p. 65] to find fixed points of T_ξ. Given an initial iterate $s_0 \in \mathbb{R}^n$, the Krasnoselskij iteration is defined as follows:

$$(10c) \qquad s_{j+1} = (1-\gamma)s_j + \gamma T_\xi(s_j),$$

with step size $\gamma \in (0,1)$. From (10c) we obtain Algorithm 1.

Algorithm 1 Low-complexity method for Problem (2)

Require: $s_0 \in \mathbb{R}^n$, $\gamma \in (0,1)$, $\epsilon_{\text{tol}} > 0$
1: **if** $z^\star := \bar{v} - R(h + H\bar{v}) \in \mathcal{Z}$ **then return** $z^\star$ ▷ check trivial solution
2: **repeat for** $j = 0, 1, \ldots$
3:      $z_{j+1} = M_\xi s_j + c_\xi$
4:      $y_{j+1} \in \mathcal{P}_\mathcal{Z}(s_j)$ ▷ projection, see Alg. 2
5:      $s_{j+1} = s_j - \gamma W(z_{j+1} - y_{j+1})$ ▷ Krasnoselskij iteration
6: **until** $\|z_{j+1} - y_{j+1}\| \leq \epsilon_{\text{tol}}$ **then return** y_{j+1} ▷ termination criterion

---

Each iteration of Algorithm 1 is simple and geared towards an efficient embedded implementation utilizing automatic code-generation. It involves only matrix-vector multiplications and a small number of projections. The projection onto the non-convex set $\mathcal{Z}$ in step 4 is the most expensive step. It can be evaluated in polynomial time using Algorithm 2, where at most $\sum_{k=1}^{N} m_k$ small-scale projections onto polyhedra are needed (Algorithm 2, step 2). State-of-the-art embedded solvers [10] or
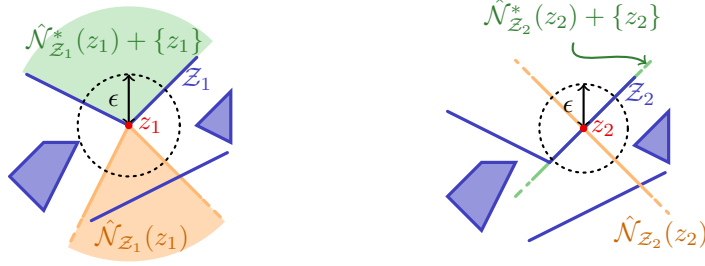
Fig. 2: Two cases of local convexifications $\hat{\mathcal{N}}_{\mathcal{Z}}^*(z) + \{z\}$ of non-convex sets $\mathcal{Z}_1$ and $\mathcal{Z}_2$ at points z_1 (left) and z_2 (right), as described in Lemma 4.1.

even explicit solutions [16] can be used for these projections. For problems of the form (3), the matrices M_ξ and W are block-banded with bandwidth $\max\{2n_x, n_u\}$, therefore steps 3 and 5 of Algorithm 1 can be evaluated very efficiently and have a computational complexity of $\mathcal{O}(N(n_x + n_u)^2)$. This enables the automatic generation of efficient, embeddable code that is tailored to particular (parametric) problems.

Algorithm 2 $\mathcal{P}_{\mathcal{Z}}(s)$, projection of s onto $\mathcal{Z}$

Require: $s = (s_1, \ldots, s_N) \in \mathbb{R}^n$, $\mathcal{Z}$ of form (1b)
 1: **for** $k = 1, \ldots, N$ **do**
 2:     **for** $i = 1, \ldots, m_k$ **do** $z_{k,i} = \mathcal{P}_{\mathcal{Z}_{k,i}}(s_k)$ ▷ small-scale projections
 3:     $z_k \in \arg\min_{i=1,\ldots,m_k} \|s_k - z_{k,i}\|$ ▷ select closest
 4: **return** $z = (z_1, \ldots, z_N)$

4.2. Local optimality. Any set $\mathcal{Z}$ of form (1b) can be transformed into a finite union of closed, convex polyhedra $\mathcal{Z}^i \subseteq \mathbb{R}^n$ with $\mathcal{Z} = \bigcup_{i=1}^I \mathcal{Z}^i$, where $I := \prod_{k=1}^N m_k$. This equivalent representation helps to simplify our theoretical considerations. The exact construction of the sets Z^i does not affect the presented results. Furthermore, we define the *set of active components* of $\mathcal{Z}$ at a point $z \in \mathcal{Z}$ as $\mathcal{I}_{\mathcal{Z}}(z) := \{i \in \{1, \ldots, I\} \mid z \in \mathcal{Z}^i\}$. To establish the subsequent theoretical results we will use the polar $\hat{\mathcal{N}}_{\mathcal{Z}}^*$ of the regular normal cone, defined in [29, Equation 6(14), p. 215] as $\hat{\mathcal{N}}_{\mathcal{Z}}^*(z) = \{v \mid \langle v, w \rangle \leq 0, \forall w \in \hat{\mathcal{N}}_{\mathcal{Z}}(z)\}$. We will show that for Problem (2) any regular KKT point corresponds to a local minimum. This implies that the fixed points of the operator T_ξ also correspond to local minima. The following lemma establishes two important properties of the set $\mathcal{Z}$ and its normal cone. These properties will be used in Theorem 4.2 for proving local optimality. At any point $z \in \mathcal{Z}$, Lemma 4.1 relates the set $\mathcal{Z}$ to a local convexification $\hat{\mathcal{N}}_{\mathcal{Z}}^*(z) + \{z\}$, as illustrated in Figure 2.

LEMMA 4.1 (Local convexification).   *For any $z \in \mathcal{Z}$, the set $\mathcal{Z}$ and the closed convex set $\hat{\mathcal{N}}_{\mathcal{Z}}^*(z) + \{z\}$ are related in the following ways:*
    i) *There exists an $\epsilon > 0$ such that $\mathcal{Z} \cap \mathcal{B}_\epsilon(z) \subseteq \hat{\mathcal{N}}_{\mathcal{Z}}^*(z) + \{z\}$.*
    ii) *Their regular normal cones at z coincide, i.e., $\hat{\mathcal{N}}_{\mathcal{Z}}(z) = \hat{\mathcal{N}}_{\hat{\mathcal{N}}_{\mathcal{Z}}^*(z)+\{z\}}(z)$.*

We now use Lemma 4.1 to prove the first key result.

THEOREM 4.2 (Sufficient local optimality condition).   *If $(z^\dagger, \lambda^\dagger)$ is a regular KKT point of Problem (2), then $z^\dagger$ is a local minimum of Problem (1).*

Proof. Let $(z^\dagger, \lambda^\dagger)$ be a regular KKT point of Problem (2). We consider the auxiliary problem:

$$
(11) \qquad \min_{z,y\,:\,z=y} \tfrac{1}{2} z^\top H z + h^\top z + \chi_{\mathcal{E}\times(\hat{\mathcal{N}}^*_{\mathcal{Z}}(z^\dagger)+\{z^\dagger\})}(z,y)\,.
$$

Problem (11) is strictly convex and its KKT conditions are given as $z = y$ (primal feasibility) and

$$
(12) \qquad 0 \in \{Hz + h\} + \hat{\mathcal{N}}_{\mathcal{E}}(z) - \{\lambda\} \text{ and } 0 \in \hat{\mathcal{N}}_{\hat{\mathcal{N}}^*_{\mathcal{Z}}(z^\dagger)+\{z^\dagger\}}(y) + \{\lambda\}\,.
$$

By Lemma 4.1(ii), (12) is equivalent to the regular KKT conditions (5). Therefore, $(z^\dagger, \lambda^\dagger)$ also satisfies (12). By strict convexity, implied by $H \succ 0$, $(z^\dagger, \lambda^\dagger)$ is primal-dual optimal for Problem (11). Furthermore, using Lemma 4.1(i), we know that there exists an $\epsilon > 0$ such that $z^\dagger \in \mathcal{Z} \cap \mathcal{B}_\epsilon(z^\dagger) \subseteq \hat{\mathcal{N}}^*_{\mathcal{Z}}(z^\dagger) + \{z^\dagger\}$, therefore, the pair $y = z = z^\dagger$ is also optimal for the non-convex problem

$$
\min_{z,y\,:\,z=y} \tfrac{1}{2} z^\top H z + h^\top z + \chi_{\mathcal{E}\times(\mathcal{Z}\cap\mathcal{B}_\epsilon(z^\dagger))}(z,y)\,,
$$

which is a local instance of Problem (2), where the feasible region is restricted to $\mathcal{B}_\epsilon(z^\dagger)$. This implies that $z^\dagger$ is locally optimal for Problem (2) and Problem (1).   □

Theorem 4.2 implies that Algorithm 1 converges to local minima.

4.3. Existence. We have seen that for any regular KKT point we can ensure the existence of a ξ-proximal KKT point by choosing ξ large enough. Therefore, if there exists a regular KKT point, we can ensure the existence of fixed points of the operator T_ξ, as demonstrated in the following theorem.

THEOREM 4.3 (Existence of fixed points of T_ξ).   *If there exists a regular KKT point $(z^\dagger, \lambda^\dagger)$ of Problem (2), then there exists a $\bar{\xi} > 0$ such that T_ξ has at least one fixed point for any $\xi \geq \bar{\xi}$.*

Proof. Given a regular KKT point $(z^\dagger, \lambda^\dagger)$ of Problem (2). By Proposition 3.2(i), there exists a $\bar{\xi} > 0$ such that for any $\xi \geq \bar{\xi}$ we have that $(z^\dagger, \lambda^\dagger)$ is a ξ-proximal KKT point. By Lemma 3.3, $s^\dagger := z^\dagger - \tfrac{1}{\xi}\lambda^\dagger \in \text{zero} \, K_\xi$ and thus $s^\dagger$ is a fixed point of T_ξ.   □

In Figure 3 we illustrate four examples for the existence and nature of KKT points. Figures 3a–3c are cases where the proximal scaling $\xi > 0$ can be chosen such that fixed points of T_ξ exist and we may find local minima using the proposed method. In Figure 3d the feasible set is a singleton and none of its corresponding KKT points are regular. In this case operator T_ξ would not have a fixed point for any $\xi > 0$, even though a generalized KKT point exists. Figures 3a and 3b illustrate common cases. On the other hand Figures 3c and 3d depict more rare cases.

4.4. Local convergence to local minima. In this section, we prove that the presented method converges locally to local minima. We say that a fixed point s° *corresponds* to a local minimum z°, if the ξ-proximal KKT point (z°, λ°) can be constructed from s° through (9). Note that, as shown in Lemma 3.3 and Theorem 4.2, any fixed point of T_ξ can be used to reconstruct a local minimum. We denote the set of fixed points corresponding to z° by $\text{fix} \, T_\xi|z^\circ := \{s \mid (z^\circ, \xi(z^\circ - s)) \text{ satisfies } (6)\}$, a convex set, see [29, p. 213]. We make the following assumptions:

(A1) Problem (2) has at least one regular KKT point.
(A2) Given any local minimum z° of Problem (1) such that $\text{fix} \, T_{\bar{\xi}}|z^\circ \neq \varnothing$ and $z^\circ \notin \text{fix} \, T_{\bar{\xi}}|z^\circ$ for some $\bar{\xi} > 0$. There exists a $\xi \geq \bar{\xi}$ and a fixed point $s^\circ \in \text{fix} \, T_\xi|z^\circ$ with multipliers $\lambda^\circ := \xi(z^\circ - s^\circ)$ such that $-\lambda^\circ \in \text{relint} \, \hat{\mathcal{N}}_{\mathcal{Z}}(z^\circ)$.
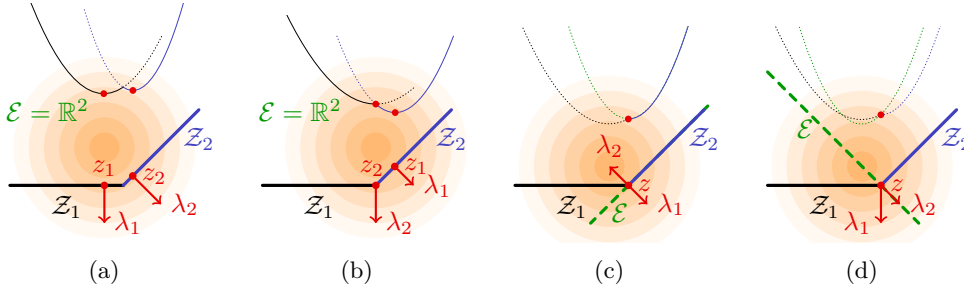
Fig. 3: KKT points of $\min_{z,y\in\mathcal{E}\times(\mathcal{Z}_1\cup\mathcal{Z}_2)\,:\,x=y}\frac{1}{2}z^\top z+h^\top z$. Level curves of the objective and the objective value evaluated on $\mathcal{Z}_1$, $\mathcal{Z}_2$ and $\mathcal{E}$ are illustrated in the respective colors, solid where feasible, dotted where infeasible. (a) (z_1,λ_1) and (z_2,λ_2) are regular KKT points corresponding to local minima. (b) (z_1,λ_1), is a regular KKT point corresponding to a local minimum, (z_2,λ_2) a non-regular KKT point corresponding to a (not locally optimal) critical point. (c) (z,λ_1), (z,λ_2) are a non-regular and a regular KKT point, corresponding to the same global minimum. (d) (z,λ_1), (z,λ_2) are the only KKT points corresponding to the unique global minimum, both are non-regular.
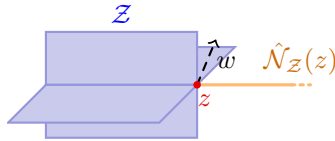

Fig. 4: A set $\mathcal{Z}$ where (A3) is violated at point z. We see that w is orthogonal to $\hat{\mathcal{N}}_{\mathcal{Z}}(z)$, but $z+\epsilon w \notin \mathcal{Z}$ for any $\epsilon > 0$.

(A3) The structure of set $\mathcal{Z}$ is such that, for any $z \in \mathcal{Z}$, either $\hat{\mathcal{N}}_{\mathcal{Z}}(z) = \{0\}$, or there exists an $\epsilon > 0$ such that for all $w \in \hat{\mathcal{N}}_{\mathcal{Z}}(z)^\perp \cap \mathcal{B}_\epsilon$, we have $z + w \in \mathcal{Z}$.

Assumption (A1) ensures the existence of regular KKT points. It furthermore implies the existence of fixed points of T_ξ for ξ large enough, according to Theorem 4.3. Assumption (A2) can be understood as a *non-degeneracy* condition on the multipliers. Assumption (A3) is an assumption on the geometry of $\mathcal{Z}$ and ensures mild local regularity. These assumptions are substantially less restrictive than common assumptions for non-convex optimization methods such as smoothness, Clarke- or prox-regularity of the constraint set $\mathcal{Z}\cap\mathcal{E}$ at critical points. In fact, we show in Section 4.5 that (A2) holds for any given Problem (1) with (almost) arbitrary convex quadratic objective and in Appendix B we present Algorithm 3 which provides a necessary and sufficient condition for (A3). Note that (A3) does not hold for all instances of Problem (1). The sets in Figure 2 and Figure 3, as well as the numerical examples in Section 5 satisfy (A3), whereas Figure 4 illustrates a simple counter-example in three dimensions.

In Algorithm 1, we apply the Krasnoselskij iteration to the operator T_ξ. This iteration is known to converge *globally* when the operator T_ξ is non-expansive, which is the case when $\mathcal{Z}$ is *convex*. To show *local* convergence for the *non-convex* case in Theorem 4.6, we need T_ξ to be non-expansive in a neighborhood around fixed points. Starting in such a neighborhood, we will converge via the same argument.
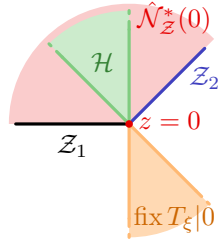
Fig. 5: The point $z = 0$ is locally optimal for the problem $\min_{z \in \mathcal{Z}} \frac{1}{2} z^\top z + h^\top z$, with $\mathcal{Z} := \mathcal{Z}_1 \cup \mathcal{Z}_2$, if and only if $h \in \mathcal{H} := -\hat{\mathcal{N}}_{\mathcal{Z}}(0)$. It is easy to verify that for all $h \in \mathcal{H}$ there exist multipliers λ such that $(0, \lambda)$ is a regular KKT point, i.e., (A1) holds. Furthermore, (A2) holds for all $\xi > \lambda_{\min}^+(R)^{-1}$ and $h \in \mathrm{relint}\,\mathcal{H}$ (almost all $h \in \mathcal{H}$), as indicated by Theorem 4.7. Also, (A3) holds for $\mathcal{Z}$. Finally, given any $h \in \mathcal{H}$: $\mathrm{fix}\,T_\xi|0 = \{-h\}$. If $h \neq 0$ ($0 \notin \mathrm{fix}\,T_\xi|0$), then $s^\circ := -h \in \mathrm{fix}\,T_\xi|0$ is unique and there exists an $\epsilon > 0$ such that the projection $\mathcal{P}_{\mathcal{Z}}(s)$ is like the projection onto the convex set $\hat{\mathcal{N}}_{\mathcal{Z}}^*(0)$ for all $s \in \mathcal{B}_\epsilon(s^\circ)$, i.e., it is single-valued, continuous and firmly non-expansive on $\mathcal{B}_\epsilon(s^\circ)$, as indicated by Lemma 4.4.

Theorem 4.6 is proved using two auxiliary lemmas. In Lemma 4.4 we show that, in a neighborhood around almost all fixed points, the projection $\mathcal{P}_{\mathcal{Z}}$ behaves like a projection onto a convex set. Therefore, it is locally firmly non-expansive, single-valued and continuous [4, Proposition 4.8, p. 61]. Then, in Lemma 4.5, we use this property, together with the particular construction of the operator T_ξ, to show that T_ξ is non-expansive, single-valued and continuous in a neighborhood around almost all fixed points. This is used in Theorem 4.6 to guarantee the *local convergence* of Algorithm 1 to a *local minimum*, almost always.

LEMMA 4.4. *Let Assumptions (A1)–(A3) hold. For any ξ large enough, and any local minimum z° to Problem (1) with $z^\circ \notin \mathrm{fix}\,T_\xi$ and $\mathrm{fix}\,T_\xi|z^\circ \neq \varnothing$; we have that for all fixed points $s^\circ \in \mathrm{fix}\,T_\xi|z^\circ$, except a measure zero subset, there exists an $\epsilon > 0$ such that the projection $\mathcal{P}_{\mathcal{Z}}$ is single-valued, continuous and firmly non-expansive on $\mathcal{B}_\epsilon(s^\circ)$.*

Note that Lemma 4.4 is weaker than prox-regularity of $\mathcal{Z}$. Prox-regularity of $\mathcal{Z}$ is equivalent to $\mathcal{P}_{\mathcal{Z}}$ being single-valued, Lipschitz-continuous and monotone in a neighborhood around any point $z \in \mathcal{Z}$, see [28, Theorem 1.3 (i)–(k), p. 5234]. Lemma 4.4 makes a similar statement about $\mathcal{P}_{\mathcal{Z}}$ for points in the neighborhood $\mathcal{B}_\epsilon(s^\circ)$ of $s^\circ \in \mathrm{fix}\,T_\xi|z^\circ$, with $\epsilon > 0$. Because Lemma 4.4 only holds when $z^\circ \notin \mathrm{fix}\,T_\xi$, the neighborhood $\mathcal{B}_\epsilon(s^\circ)$ can always be chosen such that it does not intersect with $\mathcal{Z}$. In fact, Figure 5 illustrates a simple example, where Assumptions (A1)–(A3), and thereby Lemma 4.4, hold for a set $\mathcal{Z}$ that is not prox-regular at $z = 0$.

LEMMA 4.5. *Let $\xi > 0$ and given any fixed point $s^\circ \in \mathrm{fix}\,T_\xi$ such that there exists an $\epsilon > 0$, where the projection $\mathcal{P}_{\mathcal{Z}}$ is single-valued, continuous and firmly non-expansive on $\mathcal{B}_\epsilon(s^\circ)$. Then the operator T_ξ is also single-valued, continuous and non-expansive on $\mathcal{B}_\epsilon(s^\circ)$.*

THEOREM 4.6. *Let Assumptions (A1)–(A3) hold. For any ξ large enough, and an initial iterate s_0 sufficiently close to a fixed point of T_ξ, Algorithm 1 converges almost always to a fixed point $s^\circ \in \mathrm{fix}\,T_\xi$. When it converges, z_j and y_j converge to*

each other and a local minimum of Problem (1).

Proof. Given $\xi > \lambda_{\min}^+(R)^{-1}$, i.e., (8a), and large enough, according to Theorem 4.3, (A1) implies the existence of fixed points of T_ξ, i.e., $\mathrm{fix}\, T_\xi \neq \varnothing$. We first consider that there exists a fixed point $s^\star \in \mathrm{fix}\, T_\xi$ with corresponding local minimum $z^\star$ such that $s^\star = z^\star$. Then $(z^\star, 0)$ is a KKT point of Problem (2). Therefore

$$M_\xi z^\star + c_\xi \in \mathcal{P}_\mathcal{Z}(z^\star) = \{z^\star\} \Leftrightarrow z^\star = (I - M_\xi)^{-1} c_\xi = \bar{v} - R(h + H\bar{v}),$$

with $I - M_\xi$ invertible due to $\xi > \lambda_{\min}^+(R)^{-1}$ and (8b). This is checked in step 1 of Algorithm 1, if it holds the algorithm terminates and returns $z^\star$, a global minimum. If there is no $s^\circ \in \mathrm{fix}\, T_\xi$ such that the corresponding local minimum z° satisfies $s^\circ = z^\circ$, we can apply Lemma 4.4. It says that for large enough ξ and any local minimum z°, such that $\mathrm{fix}\, T_\xi | z^\circ \neq \varnothing$, we know that for all $s^\circ \in \mathrm{fix}\, T_\xi | z^\circ$, except a measure zero subset, there exists an $\epsilon > 0$ such that the projection $\mathcal{P}_\mathcal{Z}$ is single-valued, continuous and firmly-nonexpansive on $\mathcal{B}_\epsilon(s^\circ)$. We consider any such ξ and s°. Using Lemma 4.5, T_ξ is non-expansive on $\mathcal{B}_\epsilon(s^\circ)$. If the initial iterate is close enough, i.e., $s_0 \in \mathcal{B}_\epsilon(s^\circ)$, then by the non-expansiveness of T_ξ the iterates will satisfy $s_j \in \mathcal{B}_\epsilon(s^\circ)$ for all $j \in \mathbb{N}$. The Krasnoselskij iteration applied to T_ξ, in step 5 of Algorithm 1, converges to a fixed point of T_ξ for any choice of step-size parameter $\gamma \in (0, 1)$ [6, Theorem 3.2, p. 65]. Convergence $\lim_{j \to \infty} \|z_j - y_j\| = 0$, follows by single-valuedness, and continuity of T_ξ. By Theorem 4.2, z_j, y_j converge to a local minimum of Problem (1). $\square$

Theorem 4.6 may require ξ to be large. Moreover, a lower bound of ξ satisfying Assumption (A2) may be hard to compute. However, in practice good results can be obtained for relatively small values of ξ. In addition, if the algorithm fails to converge, the theoretical results indicate that choosing a larger value of ξ may improve the convergence behavior. This is verified experimentally in Section 5.2.1, where we examine the behavior for different ξ.

4.5. Non-degeneracy. In the following we show that Assumption (A2) is satisfied for almost all instance of Problem (1). Corollary 4.8 makes the result applicable to instances of Problem (3), covering hybrid MPC.

THEOREM 4.7 (Non-degeneracy). *Consider Problem (1) with arbitrary parameter $\theta \in \mathbb{R}^p$ and cost matrix $H \succ 0$. Given any point $z \in \mathcal{E} \cap \mathcal{Z}$. For almost all linear cost terms h for which $\mathrm{fix}\, T_\xi | z \neq \varnothing$ for some $\xi > 0$, we have that (A2) holds at z.*

Proof. By Proposition 3.2(i), Lemma 3.3 and the definition of T_ξ, there exists a $\bar{\xi} > 0$ such that for all $\xi \geq \bar{\xi}$ we have that $\mathrm{fix}\, T_\xi | z \neq \varnothing$ if and only if there exists a $\lambda^\dagger$ such that $(z, \lambda^\dagger)$ is a regular KKT point. We will show that for all linear cost terms h, except a measure zero subset, there exists a λ° such that $s^\circ := z - \frac{1}{\xi} \lambda^\circ \in \mathrm{fix}\, T_\xi | z$ and $-\lambda^\circ \in \mathrm{relint}\, \hat{\mathcal{N}}_\mathcal{Z}(z)$. This means that (A2) holds at z.

The set $\mathcal{H}$ of h for which there exist $\lambda^\dagger$ such that $(z, \lambda^\dagger)$ is a regular KKT point is given by the following convex set

$$\mathcal{H} := \left\{ h \in \mathbb{R}^n \mid \exists \lambda : 0 \in \{Hz + h\} + \hat{\mathcal{N}}_\mathcal{E}(z) - \{\lambda\} \text{ and } -\lambda \in \hat{\mathcal{N}}_\mathcal{Z}(z) \right\}$$
$$= \begin{bmatrix} I & 0 & 0 \end{bmatrix} \left((\mathbb{R}^n \times -\hat{\mathcal{N}}_\mathcal{Z}(z) \times \hat{\mathcal{N}}_\mathcal{E}(z) + \{Hz\}) \cap \left\{ \begin{bmatrix} h \\ \lambda \\ v \end{bmatrix} \mid v = \lambda - h \right\} \right).$$

Using [29, p. 65–67] and convexity of the regular normal cones, we obtain

$$\text{relint}\,\mathcal{H} = \begin{bmatrix} I & 0 & 0 \end{bmatrix} \left(\left(\mathbb{R}^n \times -\text{relint}\,\hat{\mathcal{N}}_{\mathcal{Z}}(z) \times \hat{\mathcal{N}}_{\mathcal{E}}(z) + \{Hz\} \right) \cap \left\{ \begin{bmatrix} h \\ \lambda \\ v \end{bmatrix} \mid v = \lambda - h \right\} \right)$$

$$= \left\{ h \in \mathbb{R}^n \mid \exists \lambda : 0 \in \{Hz + h\} + \hat{\mathcal{N}}_{\mathcal{E}}(z) - \{\lambda\} \text{ and } -\lambda \in \text{relint}\,\hat{\mathcal{N}}_{\mathcal{Z}}(z) \right\},$$

if $\text{relint}\,\mathcal{H} \neq \varnothing$. Therefore, all linear cost terms for which (A2) holds are in relint $\mathcal{H}$. Conversely, via [21, Theorem. 1, p. 90] and convexity of $\mathcal{H}$, all linear cost terms for which (A2) does not hold are in a measure zero subset of $\mathcal{H}$. To conclude the proof, it remains to show that $\text{relint}\,\mathcal{H} \neq \varnothing$. By [29, Proposition 2.40, p. 64] and convexity of $\hat{\mathcal{N}}_{\mathcal{Z}}(z) \neq \varnothing$ we have that $\text{relint}\,\hat{\mathcal{N}}_{\mathcal{Z}}(z) \neq \varnothing$. Therefore, there exists a λ such that $-\lambda \in \text{relint}\,\hat{\mathcal{N}}_{\mathcal{Z}}(z)$. Then $h := \lambda - Hz \in \text{relint}\,\mathcal{H}$, which implies $\text{relint}\,\mathcal{H} \neq \varnothing$. □

To equivalently transform Problem (3) into Problem (1), in Remark 2.1, we have effectively restricted the choice of cost function to the case where its gradient $Hz + h$ is in the set $\mathcal{E}^\alpha$, for $\alpha \in (0,1)^N$, where the set $\mathcal{E}^\alpha$ is defined as

$$\mathcal{E}^\alpha := \bigtimes_{k=1}^{N} \left(\mathbb{R}^{n_u} \times \left\{ (w_k, x_{k+1}) \in \mathbb{R}^{2n_x} \mid \alpha_k w_k = (1 - \alpha_k) x_{k+1} \right\} \right).$$

With $\mathcal{E}$ as defined for Problem (3) in Section 2.2, we have that $\mathcal{E}^\alpha + \mathcal{E}^\perp = \mathbb{R}^n$, i.e. for any $\alpha \in (0,1)^N$: $\mathcal{E}^\alpha$ and $\mathcal{E}^\perp$ span $\mathbb{R}^n$. Theorem 4.7 does not immediately apply to this case, because the choice of cost function has been restricted. Corollary 4.8 extends Theorem 4.7 to Problem (3).

COROLLARY 4.8 (Non-degeneracy of Problem (3)). *Consider Problem (3) with arbitrary initial state $\theta \in \mathbb{R}^p$, $\alpha \in (0,1)^N$ and quadratic cost matrix $H \succ 0$ following Remark 2.1. Given any point $z \in \mathcal{E} \cap \mathcal{Z}$ and for all linear cost terms h satisfying Remark 2.1 for which $\text{fix}\,T_\xi|z \neq \varnothing$ for some $\xi > 0$, except a measure zero subset of $\mathcal{E}^\alpha$, we have that (A2) holds at z.*

Proof sketch. Following Section 2.2, any instance of Problem (3) can be written as an instance of Problem (1), where, according to Remark 2.1, we require cost functions such that $Hz + h \in \mathcal{E}^\alpha$, for some $\alpha \in (0,1)^N$ and any z. Following the proof of Theorem 4.7, we only need to show that $\text{relint}\,\mathcal{H} \cap \mathcal{E}^\alpha \neq \varnothing$. This then implies that $\text{relint}(\mathcal{H} \cap \mathcal{E}^\alpha) = \text{relint}\,\mathcal{H} \cap \mathcal{E}^\alpha$, which implies the result. As in the proof of Theorem 4.7, we have that $\text{relint}\,\hat{\mathcal{N}}_{\mathcal{Z}}(z) \neq \varnothing$. We need to show that there exists a $h \in \mathcal{E}^\alpha$ such that $(\{Hz + h\} + \hat{\mathcal{N}}_{\mathcal{E}}(z)) \cap -\text{relint}\,\hat{\mathcal{N}}_{\mathcal{Z}}(z) \neq \varnothing$. From [29, Theorem 6.46, p. 231] we have that $\hat{\mathcal{N}}_{\mathcal{E}}(z) = \mathcal{E}^\perp$, furthermore we have $Hz + \mathcal{E}^\alpha = \mathcal{E}^\alpha$, for any z, by definition. Therefore, we have $(\{Hz\} + \mathcal{E}^\alpha + \hat{\mathcal{N}}_{\mathcal{E}}(z)) \cap -\text{relint}\,\hat{\mathcal{N}}_{\mathcal{Z}}(z) = \mathbb{R}^n \cap -\text{relint}\,\hat{\mathcal{N}}_{\mathcal{Z}}(z) \neq \varnothing$.□

5. Numerical results. We consider two numerical examples and compare our method with an MIP reformulation based on disjunctions [33, Section 5], solved using CPLEX, Gurobi and MOSEK. We used YALMIP [24] as optimization interface. We further compare to an efficient implementation of non-convex ADMM, similar to [32], where the splitting (2) was used. Efficient C-code for the proposed method and ADMM was generated automatically using the developed code generation tool [11], integrating explicit solutions of the individual convex projections using MPT3 [16]. For Example 5.1 the resulting binaries are less than 75 kB. For Example 5.2 they are 490 kB for prediction horizon $N = 10$ and 550 kB for $N = 20$. The binaries for ADMM are of comparable size. In contrast, the executables for the MIP solvers are in the range of 10 to 20 MB. The reported timings are obtained on an Intel Core i7 processor *using a single core* running at 2.8 GHz. No warm-starting or re-starting is used and we always start with the initial iterate $s_0 = 0$ ($y_0 = \lambda_0 = 0$ for ADMM).

We have used step size parameter $\gamma = 0.5$ and consensus tolerance $\epsilon_{\text{tol}} = 10^{-3}$. The same tolerance was also used for the termination criterion of ADMM, according to [7], as absolute, primal and dual residual tolerance. Both examples are instances of Problem (3). They therefore satisfy (A2) for almost any linear cost term, according to Corollary 4.8. Furthermore, Assumption (A3) was verified to hold for both examples using Algorithm 3 in Appendix B.

5.1. Simple hybrid MPC. We consider a simple example taken from the hybrid MPC literature [5, Example 4.1, p. 415]. It consists of two states, one input and PWA dynamics defined over two regions:

$$x_{k+1} = \begin{cases} A_1 x_k + B u_k & \text{if } [\,1\ 0\,]x_k \geq 0 \\ A_2 x_k + B u_k & \text{if } [\,1\ 0\,]x_k \leq 0 \end{cases},$$

with $u_k \in [-1, 1]$, a regulation objective $\sum_{k=1}^{N} \|x_{k+1}\|^2 + \|u_k\|^2$ and

$$A_1 := \frac{2}{5}\begin{bmatrix} 1 & -\sqrt{3} \\ \sqrt{3} & 1 \end{bmatrix}, \ A_2 := \frac{2}{5}\begin{bmatrix} 1 & \sqrt{3} \\ -\sqrt{3} & 1 \end{bmatrix}, \ B := \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

The problem is solved for $\xi = 10$ and prediction horizons $N = 5, 10, 20, \ldots, 60$. The penalty parameter for ADMM is $\rho = 10$ and the tolerance $\epsilon_{\text{tol}} = 10^{-3}$ is used. Furthermore, a time limit of $t_{\max} = 1800\,\text{s}$ (30 minutes) is imposed for all solvers. The convergence of our method is illustrated in Figure 6a by a decreasing consensus violation $\|z_j - y_j\|$, for $N = 40$ and the first time step of the receding horizon problem. Even though not shown, the method also converges for all successive time steps (and horizons N), with similar convergence characteristics. This leads to a stable closed-loop trajectory, illustrated in Figure 6b. The trajectory is almost indistinguishable from the optimum, obtained using a MIP reformulation based on disjunctions [33, Section 5]. The computational advantage of our method is underlined by Figure 6c, where the average runtime for different control horizons N is shown. Our method is several orders of magnitude faster than the considered commercial MIP solvers. This difference becomes more pronounced for larger prediction horizons N. In fact, Gurobi (for $N \geq 30$) and CPLEX (for $N \geq 50$) frequently exceed the time limit. Furthermore, Figure 6d illustrates, that our method is slightly faster than ADMM in terms of the median runtime. The average runtime of ADMM in Figure 6c is substantially higher than for our method, because ADMM fails to converge for two of the ten time steps and terminates when it reaches 10'000 iterations. We were not able to achieve convergence of ADMM by adjusting ρ for these cases.

5.1.1. Local convergence to local minima. In order to illustrate the local convergence behavior of Algorithm 1, we have solved Example 5.1 for $N = 10$ and 50'000 different initial iterates $s_0 := z_0 - \frac{1}{\xi}\lambda_0$, where z_0 was drawn uniformly randomly from $[-1, 1]^{50}$ and λ_0 from $[-10, 10]^{50}$. The proximal scaling was chosen as $\xi = 100$ and the consensus tolerance was $\epsilon_{\text{tol}} = 10^{-8}$. Algorithm 1 converges in 99.1% of cases. In the remaining 0.9%, the initial iterate s_0 was not close enough to a fixed point (or ξ was chosen too small). The convergence to local minima can be seen in Figure 7, where we have clustered the solutions into four clusters (a)–(d) and illustrate the percentage of initial iterates achieving certain objective values which correspond to local optima. The state trajectories of the solutions for each cluster are plotted in Figures 8a–8d. Each cluster contains multiple local minima that have similar objective value. The 50'000 different initial iterates only lead to 34 trajectories that each differ
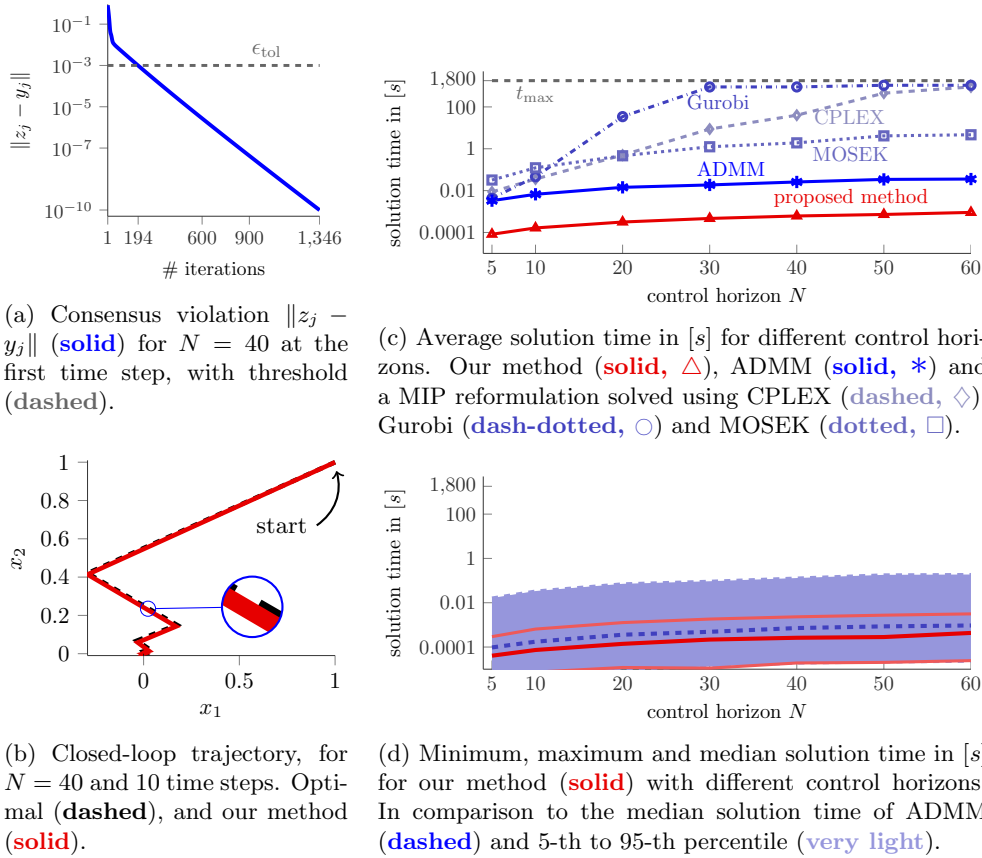
(a) Consensus violation $\|z_j - y_j\|$ (**solid**) for $N = 40$ at the first time step, with threshold (**dashed**).

(c) Average solution time in $[s]$ for different control horizons. Our method (**solid,** $\triangle$), ADMM (**solid,** $\ast$) and a MIP reformulation solved using CPLEX (**dashed,** $\diamondsuit$), Gurobi (**dash-dotted,** $\bigcirc$) and MOSEK (**dotted,** $\square$).

(b) Closed-loop trajectory, for $N = 40$ and 10 time steps. Optimal (**dashed**), and our method (**solid**).

(d) Minimum, maximum and median solution time in $[s]$ for our method (**solid**) with different control horizons. In comparison to the median solution time of ADMM (**dashed**) and 5-th to 95-th percentile (**very light**).

Fig. 6: Comparison of Example 5.1 with initial state $x_1 = x_2 = 1$.

(in terms of the objective value) by at least 10^{-5}. Cluster **(a)** with 62.1% of initial iterates contains solutions that are very close to the global optimum. The initial iterate $s_0 = 0$ used in Example 5.1 belongs to this cluster. This illustrates that the presented method is indeed a local method, and does not necessarily converge to the global optimum.

5.2. Racing. In this numerical example, we demonstrate the properties of the proposed algorithm on a more complex problem. To this end, we consider a hybrid MPC problem for racing miniature cars [23], where the friction forces acting on the tires are modeled as a PWA function, see [14, p. 108ff]. The forward velocity v_x of the car is fixed to $2 \, \mathrm{m/s}$. The state $x = (v_y, \omega)$ of the system consists of the lateral velocity v_y and the turning rate ω. The steering angle δ is the only input to the system. The continuous-time dynamics are described by

$$\dot{v}_y = \tfrac{1}{m}\left(F_{f,y}(v_y, \omega, \delta) + F_{r,y}(v_y, \omega) - m v_x \omega\right),$$
$$\dot{\omega} = \tfrac{1}{I_Z}\left(l_f F_{f,y}(v_y, \omega, \delta) - l_r F_{r,y}(v_y, \omega)\right),$$

where m, I_Z, l_f, l_r are known model parameters and $F_{f,y}$, $F_{r,y}$ are the lateral friction forces acting on the front and rear tires of the vehicle, respectively. They are given
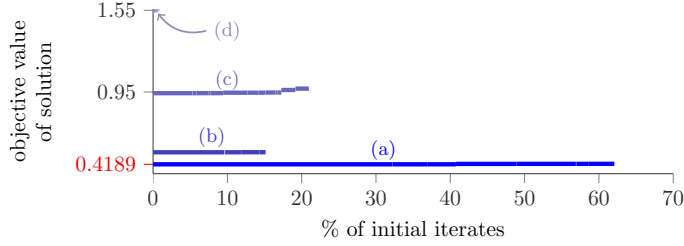
Fig. 7: Percentage of initial iterates s_0 achieving different objective values. Solutions are clustered according to objective values **(a)** in $[0.4189, 0.4225]$ (61.2%) including the optimal objective 0.4189, **(b)** in $[0.5072, 0.5078]$ (15.2%), **(c)** in $[0.9411, 0.9748]$ (21%) and **(d)** in $[1.5488, 1.5572]$ (0.9%). Each cluster contains multiple local optima.
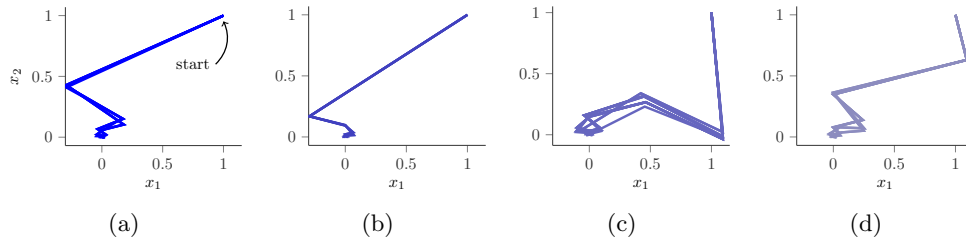

Fig. 8: Trajectories resulting from different initial iterates s_0 in clusters **(a)**–**(d)**. The number of trajectories differing (in terms of the objective) by at most 10^{-5} is **(a)** 7, **(b)** 4, **(c)** 15 and **(d)** 8.

as PWA functions with 5 pieces each, leading to a total of 19 non-empty regions per time step. For a prediction horizon N, this leads to 19^N possible combinations overall. The dynamics are discretized with a sampling time of $T_s = 20\,\text{ms}$. Additionally we impose input constraints $\delta \in [-23°, 23°]$ and state constraints:

$$(13) \qquad\qquad v_y \in [-1\,\text{m/s}, 1\,\text{m/s}], \quad \omega \in [-8\,\text{rad/s}, 8\,\text{rad/s}].$$

The objective $\sum_{k=1}^{N} \|\operatorname{diag}(1, \sqrt{10})(x_{k+1} - \bar{x}_{k+1})\|^2 + \|\delta_k - \bar{\delta}_k\|^2$, tracks a state $\bar{x}$ and input $\bar{\delta}$ reference trajectory producing an S-shaped motion of the race car.

We consider closed-loop behavior, where the MPC is applied in a receding-horizon fashion. For every time step a problem with different initial state and reference is solved and only the first input u_1 is applied to the dynamical system. A new problem with updated initial state and reference is solved for the next time step. We compare the closed-loop evolution for the different methods over 150 steps. The initial state is $v_y = \omega = 0$. The dynamics of the system are with respect to the lateral and angular velocities, v_y and ω. The MPC problem instances are solved for a prediction horizon of $N = 10$ ($\xi = 300$) and $N = 20$ ($\xi = 400$). The penalty parameter of ADMM was chosen to be $\rho = 350$ and $\rho = 450$ for $N = 10$ and $N = 20$, respectively. A time limit of 3 s for $N = 10$, and 85 s for $N = 20$ was imposed on all solvers. Both time limits are well above the computation times achieved by our algorithm, with a median runtime of 30 ms for $N = 10$, and 87 ms for $N = 20$, as reported in Figure 10. For $N = 20$, Gurobi has a median runtime of 8.5 s, reaching the time limit in some

N	our method	ADMM	Gurobi	CPLEX	MOSEK
10	1.2%	1%	3.1%	14%	92%
20	1%	0.2%	5%	21%	34%

Table 1: Comparison of relative 2-norm distance $\frac{\|z - z^\star\|}{\|z^\star\|}$ to the optimal closed-loop position trajectory $z^\star$ for $N = 10$ and $N = 20$.

cases, while CPLEX and MOSEK reach the time limit for many time steps and have a median runtime close to 85 s. The runtime of ADMM is similar to the proposed method (median runtime of 54 ms for $N = 20$), however, it fails to converge in a few cases, as shown in Figure 10. In contrast, our method converges in at most 1878 iterations for each of the 150 problems, solved in the closed-loop simulation. This is illustrated in Figure 9.
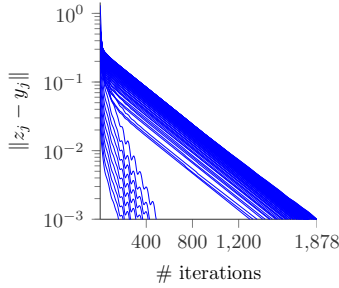

Fig. 9: Consensus violations $\|z_j - y_j\|$ for $N = 20$ and time steps 1 to 150, illustrating convergence.
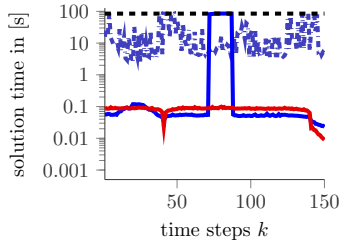

Fig. 10: Solution time in [s] for time steps k of the closed-loop simulation, for $N = 20$, time limit 85 s (**dashed**). Our method (**solid**), ADMM (**solid**) and Gurobi (**dash-dotted**).
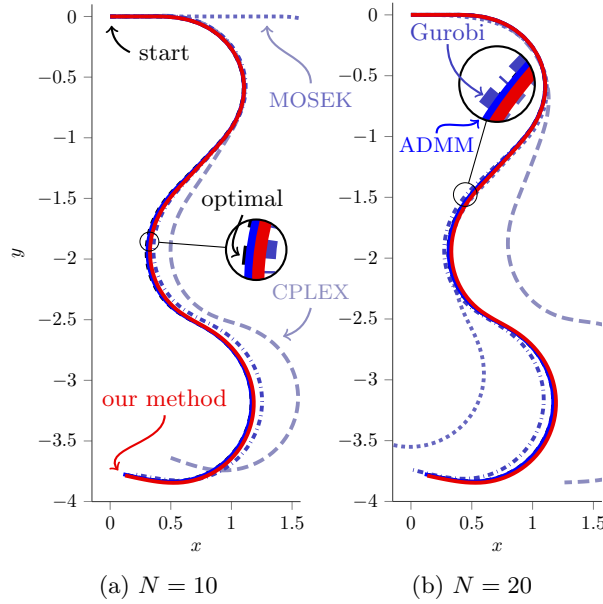

(a) $N = 10$

(b) $N = 20$

Fig. 11: Closed-loop position trajectory with initial condition $x = y = \varphi = 0$. Optimal closed-loop trajectory (**dashed**), our method (**solid**), ADMM (**solid**), CPLEX (**dashed**), Gurobi (**dash-dotted**) and MOSEK (**dotted**).

In Figure 11, we compare the position of the car, using the *position* dynamics

to transform the velocities into positions, via integration of $\dot{\varphi} = \omega$, $\dot{x} = v_x \cos(\varphi) - v_y \sin(\varphi)$ and $\dot{y} = v_x \sin(\varphi) + v_y \cos(\varphi)$. The proposed method (**solid**) visibly outperforms the commercial MIP solvers in terms of solution quality. This is also illustrated in Table 1, where the relative distance to the optimal closed-loop trajectory is reported. Our method comes very close to the optimal trajectory, while requiring several orders of magnitude less runtime. Both our method and ADMM provide near optimal solutions for similar computational effort. However, our method additionally provides guarantees on the convergence properties of the algorithm.

5.2.1. Effect of proximal scaling. To better understand the proximal scaling ξ, we consider 2000 feasible, random instances with different initial states (v_y, ω) satisfying (13). We solve the MPC problem for 40 values of ξ, with a horizon of $N = 10$ and step size $\gamma = 0.95$. Figure 12 demonstrates the relationship between ξ, (i) the number of solved instances and (ii) the number of iterations needed for convergence. For larger ξ more problems are solved, with only 4 unsolved problems remaining for $\xi = 5000$. Furthermore, the number of iterations needed to reach the consensus threshold (right axis) grows only modestly with larger ξ. In all instances, the constraint violations and relative suboptimality do not change significantly. These findings are consistent with our expectations from the theory, i.e., that larger ξ usually lead to more problems solved but at the expense of slower convergence.
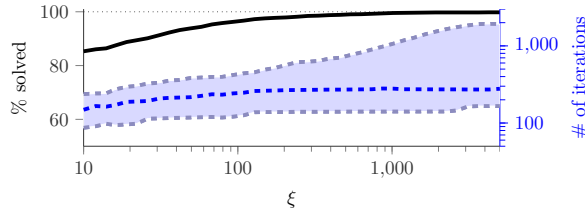

Fig. 12: Effect of proximal scaling ξ. Percentage of problems solved (**solid**, left axis) and number of iterations to reach 10^{-3} consensus violation (right axis) for the solved problems, median (**dashed**), 5-th and 95-th percentile (**dashed**).

6. Conclusion. We propose a low-complexity method for finding local minima of non-convex, non-smooth optimization problems. This simple and fast method is ideal for hybrid MPC on embedded platforms. In numerical experiments, we observe that our method provides "good" locally optimal solutions at a fraction of the time needed by global solvers. Moreover, it is competitive with ADMM in terms of speed. In contrast to other local methods, our algorithm additionally provides local optimality and local convergence guarantees.

Appendix A. Auxiliary propositions & proofs. We introduce the following additional notation: The distance of a point x to a closed set $\mathcal{C}$ is $\text{dist}(x, \mathcal{C}) := \min_{z \in \mathcal{C}} \|x - z\|$. The *relative boundary* of a set $\mathcal{C}$ is $\text{relbd}\,\mathcal{C} := \text{cl}\,\mathcal{C} \setminus \text{relint}\,\mathcal{C}$, where $\text{cl}\,\mathcal{C}$ is the *closure* and $\text{relint}\,\mathcal{C}$ the *relative interior*.

We start with a proof sketch for Lemma 3.1 utilizing [34].

Proof sketch of Lemma 3.1. Problem (2) satisfies a regularity condition called a local error bound [34, Theorem 4.3, p. 952], which means the error due to a perturbation of the constraint set can be bounded in a particular way. Therefore applying [34, Proposition 4.2, p. 951] implies that Problem (2) is calm at every local

solution, a regularity property of the optimization problem. Then using the calmness and [34, Theorem 3.6, p. 950] we obtain that for every local solution z° to Problem (2) there exists a $\lambda^\circ \in \mathbb{R}^n$ such that $(z^\circ, z^\circ, \lambda^\circ)$ is a KKT point of Problem (2).   □

To prove Proposition 3.2, we first elaborate on the structure of the regular normal cone of $\mathcal{Z}$, where the following representation is used: $\mathcal{Z} = \bigcup_{i=1}^{I} \mathcal{Z}^i$. Proposition A.1 states, that the regular normal cone of $\mathcal{Z}$ at z can be written as the finite intersection of the normal cones of its convex polyhedral parts $\mathcal{Z}^i$, where only the sets $\mathcal{Z}^i$ that contain z (i.e. the active components of $\mathcal{Z}$ at z) need to be considered.

PROPOSITION A.1 ([15, p. 59f]). *The regular normal cone $\hat{\mathcal{N}}_{\mathcal{Z}}$ of $\mathcal{Z}$ evaluated at $z \in \mathcal{Z}$ is given by $\hat{\mathcal{N}}_{\mathcal{Z}}(z) = \bigcap_{i \in \mathcal{I}_{\mathcal{Z}}(z)} \hat{\mathcal{N}}_{\mathcal{Z}^i}(z)$.*

Proof of Proposition 3.2. We consider the two statements separately.

i) Any regular KKT point (z, λ) satisfies $-\lambda \in \hat{\mathcal{N}}_{\mathcal{Z}}(z)$. Thus, by Proposition A.1

$$(14a) \qquad -\lambda \in \hat{\mathcal{N}}_{\mathcal{Z}}(z) \Rightarrow -\lambda \in \bigcap_{i \in \mathcal{I}_{\mathcal{Z}}(z)} \hat{\mathcal{N}}_{\mathcal{Z}^i}(z).$$

For any $\xi > 0$ this is equivalent to

$$(14b) \qquad (14a) \Leftrightarrow -\tfrac{1}{\xi}\lambda \in \hat{\mathcal{N}}_{\mathcal{Z}^i}(z)\, \forall i \in \mathcal{I}_{\mathcal{Z}}(z) \Leftrightarrow z \in \mathcal{P}_{\mathcal{Z}^i}(z - \tfrac{1}{\xi}\lambda)\, \forall i \in \mathcal{I}_{\mathcal{Z}}(z),$$

where (14b) follows from the convexity of $\mathcal{Z}^i$ and the relationship between $\mathcal{P}_{\mathcal{Z}^i}$ and $\hat{\mathcal{N}}_{\mathcal{Z}^i}$, described in [29, p. 212f]. By construction, for any $\epsilon > 0$ small enough, we have $\mathcal{Z} \cap \mathcal{B}_\epsilon(z) = \bigcup_{i \in \mathcal{I}_{\mathcal{Z}}(z)} \mathcal{Z}^i \cap \mathcal{B}_\epsilon(z)$ and thus $(14b) \Rightarrow z \in \mathcal{P}_{\mathcal{Z} \cap \mathcal{B}_\epsilon(z)}(z - \tfrac{1}{\xi}\lambda)$. We consider $\bar{\xi} := \tfrac{2}{\epsilon}\|\lambda\|$. For any $\xi \geq \bar{\xi}$ we have $\|(z - \tfrac{1}{\xi}\lambda) - z\| = \tfrac{1}{\xi}\|\lambda\| \leq \tfrac{\epsilon}{2}$. Thus, for all $x \in \mathcal{Z}$ with $\|z - x\| > \epsilon$ we have $\|(z - \tfrac{1}{\xi}\lambda) - x\| > \tfrac{\epsilon}{2}$. Together with $z \in \mathcal{P}_{\mathcal{Z} \cap \mathcal{B}_\epsilon(z)}(z - \tfrac{1}{\xi}\lambda)$, this implies that $z \in \mathcal{P}_{\mathcal{Z}}(z - \tfrac{1}{\xi}\lambda)$ for all $\xi \geq \bar{\xi} = \tfrac{2}{\epsilon}\|\lambda\|$, which implies that (z, λ) is a ξ-proximal KKT point.

ii) Given any $\xi > 0$, and ξ-proximal KKT point (z, λ). By definition $z \in \mathcal{P}_{\mathcal{Z}}(z - \tfrac{1}{\xi}\lambda)$ which implies $-\tfrac{1}{\xi}\lambda \in \hat{\mathcal{N}}_{\mathcal{Z}}(z)$ via [29, Example 6.16, p. 212]. This is equivalent to $-\lambda \in \hat{\mathcal{N}}_{\mathcal{Z}}(z)$, which directly implies that (z, λ) is a regular KKT point.   □

Proof of Lemma 3.3. Given any $\xi > 0$, and any $s^\dagger \in$ zero K_ξ, by definition $0 \in \{M_\xi s^\dagger + c_\xi\} - \mathcal{P}_{\mathcal{Z}}(s^\dagger)$, and thereby $z^\dagger := M_\xi s^\dagger + c_\xi \in \mathcal{P}_{\mathcal{Z}}(s^\dagger)$. We define $\lambda^\dagger := \xi(z^\dagger - s^\dagger)$ and thus $z^\dagger \in \mathcal{P}_{\mathcal{Z}}(z^\dagger - \tfrac{1}{\xi}\lambda^\dagger) \Leftrightarrow$ (6b) holds. Furthermore, we have $z^\dagger = M_\xi(z^\dagger - \tfrac{1}{\xi}\lambda^\dagger) + c_\xi$ which, using (7b), is equivalent to $z^\dagger = R\left(\lambda^\dagger - h - H\bar{v}\right) + \bar{v}$. By strict convexity $z^\dagger$ is the unique solution to $0 \in \partial_z(\tfrac{1}{2}z^\top H z + h^\top z - \lambda^{\dagger\top} z + \chi_{\mathcal{E}}(z))(z^\dagger)$, which is equivalent to $0 \in \{Hz^\dagger + h - \lambda^\dagger\} + \hat{\mathcal{N}}_{\mathcal{E}}(z^\dagger)$ and therefore (6a) holds and $(z^\dagger, \lambda^\dagger)$ is a ξ-proximal KKT point.   □

Proof of Lemma 4.1. We prove the two parts individually.

i) Take any $\bar{z} \in \mathcal{Z}$ and $z \in \mathcal{Z} \cap \mathcal{B}_\epsilon(\bar{z})$. Using the definition of the polar cone we have that $z \in \hat{\mathcal{N}}_{\mathcal{Z}}^*(\bar{z}) + \{\bar{z}\}$ if and only if $\langle z - \bar{z}, v \rangle \leq 0$ for all $v \in \hat{\mathcal{N}}_{\mathcal{Z}}(\bar{z})$. We assume for the sake of contradiction that for all $\epsilon > 0$ there exists a $v \in \hat{\mathcal{N}}_{\mathcal{Z}}(\bar{z})$ such that $\langle z - \bar{z}, v \rangle > 0$. Using Proposition A.1, we have that for any $\epsilon > 0$ small enough $v \in \hat{\mathcal{N}}_{\mathcal{Z}}(\bar{z})$ is equivalent to $\langle v, u - \bar{z} \rangle \leq 0$ for all $u \in \mathcal{Z}^i$, and all $i \in \mathcal{I}_{\mathcal{Z}}(\bar{z})$. However, since $z \in \mathcal{Z}$, there exists an $i \in \mathcal{I}_{\mathcal{Z}}(\bar{z})$ such that $z \in \mathcal{Z}^i$, and therefore by choosing $u = z$ we have $\langle v, u - \bar{z} \rangle < \langle z - \bar{z}, v \rangle$, a contradiction.

ii) For any $z \in \mathcal{Z}$, we have that $\hat{\mathcal{N}}_{\mathcal{Z}}(z) = \hat{\mathcal{N}}_{\mathcal{Z}}^{**}(z)$, by [29, Corollary 6.21, p. 216] and $\hat{\mathcal{N}}_{\mathcal{Z}}(z)$ closed and convex. Using the definition of the polar and regular normal

cones [29, p. 203, p. 215], and convexity of $\hat{\mathcal{N}}_{\mathcal{Z}}^*(z)$ we have that $v \in \hat{\mathcal{N}}_{\mathcal{Z}}(z)$ is equivalent to $\langle v, w \rangle \leq 0$ for all $w \in \hat{\mathcal{N}}_{\mathcal{Z}}^*(z)$, which is equivalent to $\langle v, u - z \rangle \leq 0$ for all $u \in \hat{\mathcal{N}}_{\mathcal{Z}}^*(z) + \{z\}$. This means $v \in \hat{\mathcal{N}}_{\hat{\mathcal{N}}_{\mathcal{Z}}^*(z)+\{z\}}(z)$ and implies the result. □

The following proposition is an auxiliary proposition needed to prove Lemma 4.4. It gives a local characterization of the projection around points $\bar{s}$ where it is unique. It states, that there always exists a neighborhood $\mathcal{B}_\epsilon(\bar{s})$, with $\epsilon > 0$, such that for all points in the neighborhood the projection onto $\mathcal{Z}$ can be restricted to the projection onto just the union of the active components of $\mathcal{Z}$ at $\bar{z} = \mathcal{P}_{\mathcal{Z}}(\bar{s})$. Furthermore for all points $s \in \mathcal{B}_\epsilon(\bar{s})$ in the neighborhood the result $z \in \mathcal{P}_{\mathcal{Z}}(s)$ of the projection satisfies the bound $\|z - \bar{z}\| \leq \|s - \bar{s}\|$.

PROPOSITION A.2 (Projection).   *Given $\bar{z}$ and $\bar{s}$ such that $\mathcal{P}_{\mathcal{Z}}(\bar{s}) = \{\bar{z}\}$, then there exists an $\epsilon > 0$ such that for all $s \in \mathcal{B}_\epsilon(\bar{s})$: $\mathcal{P}_{\mathcal{Z}}(s) = \mathcal{P}_{\mathcal{Y}(\bar{z})}(s) \subseteq \mathcal{B}_{\|s-\bar{s}\|}(\bar{z})$, where $\mathcal{Y}(z) := \bigcup_{i \in \mathcal{I}_{\mathcal{Z}}(z)} \mathcal{Z}^i$.*

Proof. First, we will show that for all $s \in \mathbb{R}^n$, we have $\mathcal{P}_{\mathcal{Y}(\bar{z})}(s) \subseteq \mathcal{B}_{\|s-\bar{s}\|}(\bar{z})$. For all $i \in \mathcal{I}_{\mathcal{Z}}(\bar{z})$, the projection $\mathcal{P}_{\mathcal{Z}^i}$ is non-expansive because $\mathcal{Z}^i$ is closed and convex. Therefore, for any $i \in \mathcal{I}_{\mathcal{Z}}(\bar{z})$ and any $s \in \mathbb{R}^n$ we have $\|\mathcal{P}_{\mathcal{Z}^i}(s) - \mathcal{P}_{\mathcal{Z}^i}(\bar{s})\| \leq \|s - \bar{s}\|$. Consider any $z \in \mathcal{P}_{\mathcal{Y}(\bar{z})}(s)$, since $\bar{z}$ and $\bar{s}$ are such that $\mathcal{P}_{\mathcal{Z}}(\bar{s}) = \{\bar{z}\}$ it follows that $\mathcal{P}_{\mathcal{Y}(\bar{z})}(\bar{s}) = \mathcal{P}_{\mathcal{Z}^i}(\bar{s}) = \{\bar{z}\}$ for all $i \in \mathcal{I}_{\mathcal{Z}}(\bar{z})$. In particular for some $i \in \mathcal{I}_{\mathcal{Z}}(\bar{z})$ we have

$$(15) \qquad \|z - \bar{z}\| = \|\mathcal{P}_{\mathcal{Z}^i}(s) - \mathcal{P}_{\mathcal{Z}^i}(\bar{s})\| \leq \|s - \bar{s}\|,$$

which implies $\mathcal{P}_{\mathcal{Y}(\bar{z})}(s) \subseteq \mathcal{B}_{\|s-\bar{s}\|}(\bar{z})$. Second, we show that there exists an $\epsilon > 0$ such that for any $y \in \mathcal{Z} \setminus \mathcal{Y}(\bar{z})$

$$(16a) \qquad \|y - s\| > \|z - s\| \quad \forall s \in \mathcal{B}_\epsilon(\bar{s}), z \in \mathcal{P}_{\mathcal{Y}(\bar{z})}(s).$$

We define $\varsigma := \inf_{y \in \mathcal{Z} \setminus \mathcal{Y}(\bar{z})} \|y - \bar{s}\| - \|\bar{z} - \bar{s}\| > 0$, by closedness of $\mathcal{Z}$ and $\mathcal{Z}^i$. The triangle inequality implies that for any $y \in \mathcal{Z} \setminus \mathcal{Y}(\bar{z})$ we have $\|y - \bar{s}\| \leq \|y - s\| + \|s - \bar{s}\|$. We choose $\epsilon < \frac{\varsigma}{3}$, which implies

$$(16b) \quad \|y - s\| \geq \|y - \bar{s}\| - \|s - \bar{s}\| \geq \varsigma + \|\bar{z} - \bar{s}\| - \|s - \bar{s}\| > 3\epsilon + \|\bar{z} - \bar{s}\| - \epsilon,$$

where in the second step we have used the definition of ς, and in the third step we have used $\varsigma > 3\epsilon$ and $\|s - \bar{s}\| \leq \epsilon$. Now we consider any $z \in \mathcal{P}_{\mathcal{Y}(\bar{z})}(s)$, via the triangle inequality $\|z - s\| \leq \|z - \bar{z}\| + \|s - \bar{s}\| + \|\bar{z} - \bar{s}\|$ and (15) we have

$$\|z - s\| \leq 2\|s - \bar{s}\| + \|\bar{z} - \bar{s}\| \leq 2\epsilon + \|\bar{z} - \bar{s}\|,$$

which together with (16b) implies (16a). This implies the result. □

PROPOSITION A.3. *There exists a $\bar{\xi} > 0$ such that for any local optimum z°, either $\mathrm{fix}\, T_\xi | z^\circ = \varnothing$ for all $\xi \geq \bar{\xi}$ or $\mathrm{fix}\, T_\xi | z^\circ \neq \varnothing$ for all $\xi \geq \bar{\xi}$.*

Proof. Given any local optimum z°. Either $\mathrm{fix}\, T_{\bar{\xi}} | z^\circ = \varnothing$ for all $\bar{\xi} > 0$ or there exists a $\bar{\xi}$ such that there exists an $s \in \mathrm{fix}\, T_{\bar{\xi}} | z^\circ$. By Lemma 3.3, it holds that (z°, λ), with $\lambda := \bar{\xi}(z^\circ - s)$, is a $\bar{\xi}$-proximal KKT point. From Proposition 3.2, it further follows that (z°, λ) is also a ξ-proximal KKT point for any $\xi \geq \bar{\xi}$, which implies $s := z^\circ - \frac{1}{\xi}\lambda \in \mathrm{fix}\, T_\xi | z^\circ \neq \varnothing$. □

In order to prove Lemma 4.4, we make the following claim.

Claim A.4. Given Assumptions (A1)–(A2). For any $\xi > 0$ large enough and any local minimum z° to Problem (2), with fix $T_\xi | z^\circ \neq \varnothing$ and $z^\circ \notin \text{fix } T_\xi | z^\circ$, we have that

$$(17) \qquad \mathcal{P}_{\mathcal{Z}}(s^\circ) = \{z^\circ\} \text{ and } s^\circ \in \text{relint } \hat{\mathcal{N}}_{\mathcal{Z}}(z^\circ) + \{z^\circ\},$$

for all fixed points $s^\circ \in \text{fix } T_\xi | z^\circ$, except a measure zero subset.

The proof of Claim A.4 comes directly after the proof of Lemma 4.4.

Proof of Lemma 4.4. Given Claim A.4 and consider any pair z°, s° satisfying (17). We will show that in a neighborhood of s° the projection onto $\mathcal{Z}$ is the same as projecting onto an appropriate closed, convex set. This then implies local firm non-expansiveness, continuity and single-valuedness of the projection.

Let $\mathcal{B}_\epsilon(s^\circ)$ be a neighborhood of s°, for some $\epsilon > 0$ small enough. We define $\mathcal{S} := \text{lin}(\hat{\mathcal{N}}_{\mathcal{Z}}(z^\circ))$, the linear subspace spanned by $\hat{\mathcal{N}}_{\mathcal{Z}}(z^\circ)$, and its orthogonal complement $\mathcal{S}^\perp$. For any $s \in \mathcal{B}_\epsilon(s^\circ)$, we can write $s = s^\circ + u + v$, where $u \in \mathcal{S}$, $v \in \mathcal{S}^\perp$ and $u + v \in \mathcal{B}_\epsilon$. This allows us to argue about the contribution of u and v successively. We consider $\bar{s} := s^\circ + u$. Due to $\mathcal{P}_{\mathcal{Z}}(s^\circ) = \{z^\circ\}$ and [29, Example 6.16, p. 212], we have $\mathcal{P}_{\mathcal{Z}}(\bar{s}) = \mathcal{P}_{\mathcal{Z}}(s^\circ + u) = \{z^\circ\}$ for all $u \in \mathcal{B}_\epsilon$. Furthermore, due to openness of $\text{relint } \hat{\mathcal{N}}_{\mathcal{Z}}(z^\circ)$ on $\mathcal{S}$, we have $\bar{s} \in \text{relint } \hat{\mathcal{N}}_{\mathcal{Z}}(z^\circ) + \{z^\circ\}$ for all $u \in \mathcal{B}_\epsilon$, with $\epsilon > 0$ small enough. Therefore instead of considering a pair z°, s° satisfying (17), we will consider $z^\circ, \bar{s}$. We now examine the contribution of v, i.e., $s = \bar{s} + v$. Similarly to Lemma 4.1, where we showed that we can replace the regular normal cone of $\mathcal{Z}$ by the normal cone of $\hat{\mathcal{N}}_{\mathcal{Z}}^*(z^\circ) + \{z\}$, we will now consider the projection onto that set. We do this because in some sense $\hat{\mathcal{N}}_{\mathcal{Z}}^*(z^\circ) + \{z^\circ\}$ is a local convexification of $\mathcal{Z}$. We have $v \in \mathcal{S}^\perp \subseteq \hat{\mathcal{N}}_{\mathcal{Z}}^*(z^\circ)$. We consider the projection $\mathcal{P}_{\hat{\mathcal{N}}_{\mathcal{Z}}^*(z^\circ)}(s - z^\circ) = \arg\min_{w \in \hat{\mathcal{N}}_{\mathcal{Z}}^*(z^\circ)} \|\bar{s} - z^\circ + v - w\|^2 = \arg\min_{w \in \hat{\mathcal{N}}_{\mathcal{Z}}^*(z^\circ)} \|\bar{s} - z^\circ\|^2 + \|v - w\|^2 - 2\langle \bar{s} - z^\circ, w\rangle$. Where we have used $\langle \bar{s} - z^\circ, v\rangle = 0$. By $\langle \bar{s} - z^\circ, w\rangle \leq 0$, v is a minimizer. It is unique because the set $\hat{\mathcal{N}}_{\mathcal{Z}}^*(z^\circ)$ is closed and convex, therefore $\mathcal{P}_{\hat{\mathcal{N}}_{\mathcal{Z}}^*(z^\circ)}(s - z^\circ) = \{v\}$ and consequently $\mathcal{P}_{\hat{\mathcal{N}}_{\mathcal{Z}}^*(z^\circ) + \{z^\circ\}}(\bar{s} + v) = \{z^\circ + v\}$. Additionally, due to $\mathcal{Z} \cap \mathcal{B}_\epsilon(z^\circ) \subseteq \hat{\mathcal{N}}_{\mathcal{Z}}^*(z^\circ) + \{z^\circ\}$ from Lemma 4.1 and $z^\circ + v \in \mathcal{Z} \cap \mathcal{B}_\epsilon(z^\circ)$ due to Assumption (A3), it follows that $z^\circ + v \in \mathcal{P}_{\mathcal{Z}}(\bar{s} + v)$. Finally using Proposition A.2 we know that $\mathcal{P}_{\mathcal{Z}}(\bar{s} + v) \subseteq \mathcal{Z} \cap \mathcal{B}_{\|v\|}(z^\circ) \subseteq \mathcal{Z} \cap \mathcal{B}_\epsilon(z^\circ)$, which immediately implies that $\mathcal{P}_{\mathcal{Z}}(\bar{s} + v) = \mathcal{P}_{\hat{\mathcal{N}}_{\mathcal{Z}}^*(z^\circ) + \{z^\circ\}}(\bar{s} + v) = \{z^\circ + v\}$. This means for all $s \in \mathcal{B}_\epsilon(s^\circ)$ the projection onto $\mathcal{Z}$ is equivalent to the projection onto the closed, convex set $\hat{\mathcal{N}}_{\mathcal{Z}}^*(z^\circ) + \{z^\circ\}$ and therefore enjoys all of the properties of a projection onto a closed convex set. In particular single-valuedness, continuity and firm non-expansiveness [4, Proposition 4.8, p. 61]. □

Proof of Claim A.4. Given a local minimum z° to Problem (2) and consider any $\bar{\xi}$ according to Proposition A.3 such that (A2) holds. By definition of fix T_ξ and by [29, Example 6.16, p. 212], (A2) also holds for any larger $\xi \geq \bar{\xi}$. We will show that for any $\xi > \bar{\xi}$, and any local minimum z° such that fix $T_\xi | z^\circ \neq \varnothing$ and $z^\circ \notin \text{fix } T_\xi | z^\circ$:

$$\text{relint fix } T_\xi | z^\circ = \left\{ s \in \text{fix } T_\xi | z^\circ \mid \mathcal{P}_{\mathcal{Z}}(s) = \{z^\circ\}, \ s \in \text{relint } \hat{\mathcal{N}}_{\mathcal{Z}}(z^\circ) + \{z^\circ\}\right\}.$$

By convexity of fix $T_\xi | z^\circ$ and [21, Theorem. 1, p. 90], this implies that (17) holds for all $s \in \text{fix } T_\xi | z^\circ$, except a measure zero subset.

Due to fix $T_{\bar{\xi}} | z^\circ \neq \varnothing$ and (A2), we have that there exists a $\bar{s} \in \text{fix } T_{\bar{\xi}} | z^\circ$ with $\bar{s} \in \text{relint } \hat{\mathcal{N}}_{\mathcal{Z}}(z^\circ) + \{z^\circ\}$. Furthermore, we have that $s(\xi) := z^\circ - \frac{1}{\xi}\bar{\xi}(z^\circ - \bar{s}) \in \text{fix } T_\xi | z^\circ$ and $s(\xi) \in \text{relint } \hat{\mathcal{N}}_{\mathcal{Z}}(z^\circ) + \{z^\circ\}$ for $\xi \geq \bar{\xi}$ and by [29, Example 6.16, p. 212] we have

that $\mathcal{P}_{\mathcal{Z}}(s(\xi)) = \{z^\circ\}$ for any $\xi > \bar{\xi}$. Therefore, for any $\xi > \bar{\xi}$ we have that

$$(18) \qquad \text{fix } T_\xi | z^\circ \cap \text{relint} \left(\hat{\mathcal{N}}_{\mathcal{Z}}(z^\circ) + \{z^\circ\} \right) \cap \{ s \mid \mathcal{P}_{\mathcal{Z}}(s) = \{z^\circ\} \} \neq \varnothing \,.$$

We consider the set $\{ s \mid z^\circ \in \mathcal{P}_{\mathcal{Z}}(s) \}$. We have

$$
\begin{aligned}
\{ s \mid z^\circ \in \mathcal{P}_{\mathcal{Z}}(s) \} &= \{ s \mid \|s - z^\circ\| \leq \text{dist}(s, \mathcal{Z}) \} \\
&= \{ s \mid \|s - z^\circ\| \leq \text{dist}(s, \mathcal{Y}(z^\circ)) \} \cap \{ s \mid \|s - z^\circ\| \leq \text{dist}(s, \mathcal{Y}^c(z^\circ)) \} \\
&= \left(\hat{\mathcal{N}}_{\mathcal{Z}}(z^\circ) + \{z^\circ\} \right) \cap \{ s \mid \|s - z^\circ\| \leq \text{dist}(s, \mathcal{Y}^c(z^\circ)) \} \,,
\end{aligned}
$$

with $\mathcal{Y}(z^\circ) := \bigcup_{i \in \mathcal{I}_{\mathcal{Z}}(z^\circ)} \mathcal{Z}^i$ and $\mathcal{Y}^c(z^\circ) := \bigcup_{i \in \mathcal{I}_{\mathcal{Z}}^c(z^\circ)} \mathcal{Z}^i$, where $\mathcal{I}_{\mathcal{Z}}^c(z^\circ) := \{1, \dots, I\} \setminus \mathcal{I}_{\mathcal{Z}}(z^\circ)$. Where we have used that by Proposition A.1 and convexity of $\mathcal{Z}^i$ we have

$$
\begin{aligned}
\{ s \mid \|s - z^\circ\| \leq \text{dist}(s, \mathcal{Y}(z^\circ)) \} &= \{ s \mid z^\circ \in \mathcal{P}_{\mathcal{Y}(z^\circ)}(s) \} \\
&= \hat{\mathcal{N}}_{\mathcal{Y}(z^\circ)}(z^\circ) + \{z^\circ\} = \hat{\mathcal{N}}_{\mathcal{Z}}(z^\circ) + \{z^\circ\} \,.
\end{aligned}
$$

By continuity and upper-boundedness of $\varsigma(s) := \|s - z^\circ\| - \text{dist}(s, \mathcal{Y}^c(z^\circ))$, and by closedness of $\mathcal{Y}^c(z^\circ)$, it follows that

$$
\begin{aligned}
\text{relint} \left\{ s \mid \|s - z^\circ\| \leq \text{dist}(s, \mathcal{Y}^c(z^\circ)) \right\} &= \{ s \mid \|s - z^\circ\| < \text{dist}(s, \mathcal{Y}^c(z^\circ)) \} \\
&= \{ s \mid \mathcal{P}_{\mathcal{Z}}(s) = \{z^\circ\} \}
\end{aligned}
$$

for all s such that $z^\circ \in \mathcal{P}_{\mathcal{Z}}(s)$. Therefore, by [29, Proposition 2.42, p. 65] and (18)

$$\text{relint} \left\{ s \mid z^\circ \in \mathcal{P}_{\mathcal{Z}}(s) \right\} = \text{relint} \left(\hat{\mathcal{N}}_{\mathcal{Z}}(z^\circ) + \{z^\circ\} \right) \cap \{ s \mid \mathcal{P}_{\mathcal{Z}}(s) = \{z^\circ\} \} \neq \varnothing \,.$$

By definition of fix $T_\xi | z^\circ$, convexity of $\{ s \mid z^\circ \in \mathcal{P}_{\mathcal{Z}}(s) \}$, [29, Proposition 2.42, p. 65] and (18) we then have

$$
\begin{aligned}
\text{relint fix } T_\xi | z^\circ &= \text{relint} \left\{ s \mid 0 \in \{ H z^\circ + h \} + \hat{\mathcal{N}}_{\mathcal{E}}(z^\circ) - \{ \xi(z^\circ - s) \}, \ z^\circ \in \mathcal{P}_{\mathcal{Z}}(s) \right\} \\
&= \text{fix } T_\xi | z^\circ \cap \text{relint} \left\{ s \mid z^\circ \in \mathcal{P}_{\mathcal{Z}}(s) \right\} \\
&= \text{fix } T_\xi | z^\circ \cap \text{relint} \left(\hat{\mathcal{N}}_{\mathcal{Z}}(z^\circ) + \{z^\circ\} \right) \cap \{ s \mid \mathcal{P}_{\mathcal{Z}}(s) = \{z^\circ\} \} \neq \varnothing \,. \qquad \square
\end{aligned}
$$

Proof of Lemma 4.5. We consider any $\xi > 0$, $\epsilon > 0$ and $s^\circ \in \text{fix } T_\xi$ such that the projection $\mathcal{P}_{\mathcal{Z}}$ is single-valued, continuous and firmly non-expansive on $\mathcal{B}_\epsilon(s^\circ)$. We recall the definition of T_ξ in (10a). Clearly T_ξ is single-valued and continuous on $\mathcal{B}_\epsilon(s^\circ)$ by virtue of the projection. It remains to show the non-expansiveness of T_ξ on $\mathcal{B}_\epsilon(s^\circ)$. By [4, Definition 4.1 (ii), p. 59], T_ξ is non-expansive on $\mathcal{B}_\epsilon(s^\circ)$ if $\|T_\xi(s) - T_\xi(t)\|^2 \leq \|s - t\|^2$ for all $s, t \in \mathcal{B}_\epsilon(s^\circ)$. For any s, t, we have that

$$
\begin{aligned}
(19) \qquad \|T_\xi(s) - T_\xi(t)\|^2 &= \|(I - W M_\xi)(s - t)\|^2 \\
&\quad + 2\langle W^\top (I - W M_\xi)(s - t), \mathcal{P}_{\mathcal{Z}}(s) - \mathcal{P}_{\mathcal{Z}}(t) \rangle \\
&\quad + \|W(\mathcal{P}_{\mathcal{Z}}(s) - \mathcal{P}_{\mathcal{Z}}(t))\|^2 \,.
\end{aligned}
$$

To show $(19) \leq \|s - t\|^2$ for all $s, t \in \mathcal{B}_\epsilon(s^\circ)$ we will split (19) into two parts, the portions belonging to the range- and nullspace of M_ξ. We recall that M_ξ is positive semi-definite for ξ satisfying (8a). Therefore, we can consider the eigenvalue decomposition of M_ξ, as in (10) and define q_i, for $i = 1, \dots, n$, to be the normalized eigenvectors of M_ξ. We define $\delta, p \in \mathbb{R}^n$ as coefficients, such that $\sum_{i=1}^n \delta_i q_i = s - t$ and $\sum_{i=1}^n p_i q_i =$

$\mathcal{P}_{\mathcal{Z}}(s) - \mathcal{P}_{\mathcal{Z}}(t)$. We further define $\delta^{\perp} := \sum_{i=1}^{n-m} \delta_i q_i$, $\delta^0 := \sum_{i=n-m+1}^{n} \delta_i q_i$, $p^{\perp} := \sum_{i=1}^{n-m} p_i q_i$ and $p^0 := \sum_{i=n-m+1}^{n} p_i q_i$, which splits $s - t$ and $\mathcal{P}_{\mathcal{Z}}(s) - \mathcal{P}_{\mathcal{Z}}(t)$ into their range and nullspace portions. Additionally, $p_{\Lambda}^{\perp} := \sum_{i=1}^{n-m} \lambda_i^{-1} p_i q_i$, for λ_i the non-zero eigenvalues of M_{ξ}. Substituting $\delta^{\perp}$, δ^0, $p^{\perp}$, p^0 and $p_{\Lambda}^{\perp}$ into (19), using the eigenvalue decomposition of M_{ξ} and W as defined in (10b), we obtain

$$(20) \quad \|T_{\xi}(s) - T_{\xi}(t)\|^2 = \tfrac{1}{4}\|\delta^{\perp}\|^2 + \|\delta^0\|^2 + \tfrac{1}{2}\langle \delta^{\perp}, p_{\Lambda}^{\perp} \rangle - 2\langle \delta^0, p^0 \rangle + \tfrac{1}{4}\|p_{\Lambda}^{\perp}\|^2 + \|p^0\|^2 .$$

We want to show that $(19) = (20) \leq \|s - t\|^2 = \|\delta^{\perp}\|^2 + \|\delta^0\|^2$. However, this only has to hold for δ, p that are related via the projection $\mathcal{P}_{\mathcal{Z}}$ where in particular we want to exploit its firm non-expansiveness. The projection $\mathcal{P}_{\mathcal{Z}}$ being firmly non-expansive on $\mathcal{B}_{\epsilon}(s^{\circ})$, means by [4, Proposition 4.2 (iv), p. 60], that for all $s, t \in \mathcal{B}_{\epsilon}(s^{\circ})$

$$(\text{FNE}) \qquad \|\mathcal{P}_{\mathcal{Z}}(s) - \mathcal{P}_{\mathcal{Z}}(t)\|^2 \leq \langle s - t, \mathcal{P}_{\mathcal{Z}}(s) - \mathcal{P}_{\mathcal{Z}}(t) \rangle .$$

It is also non-expansive, i.e., for any $s, t \in \mathcal{B}_{\epsilon}(s^{\circ})$

$$(\text{NE}) \qquad \|\mathcal{P}_{\mathcal{Z}}(s) - \mathcal{P}_{\mathcal{Z}}(t)\|^2 \leq \|s - t\|^2 ,$$

By applying the Cauchy-Schwarz inequality, and combining (FNE) and (NE) we get

$$(21) \qquad 0 \leq \|\mathcal{P}_{\mathcal{Z}}(s) - \mathcal{P}_{\mathcal{Z}}(t)\|^2 \leq \langle s - t, \mathcal{P}_{\mathcal{Z}}(s) - \mathcal{P}_{\mathcal{Z}}(t) \rangle \leq \|s - t\|^2 .$$

Substituting $\delta^{\perp}, \delta^0, p^{\perp}$ and p^0 into (21), we obtain

$$(22) \qquad 0 \leq \|p^{\perp}\|^2 + \|p^0\|^2 \leq \langle p^{\perp}, \delta^{\perp} \rangle + \langle p^0, \delta^0 \rangle \leq \|\delta^{\perp}\|^2 + \|\delta^0\|^2 .$$

To show that T_{ξ} is non-expansive it is therefore sufficient to show that for all δ, p that satisfy (22) it holds that $(20) \leq \|\delta^{\perp}\|^2 + \|\delta^0\|^2$. Notice, that (20) and (22) do not contain any coupling terms between range- and nullspace portions of δ or p, therefore we can argue about them separately.

i) We consider all $\delta^{\perp}, p^{\perp}$ such that $0 \leq \|p^{\perp}\|^2 \leq \langle p^{\perp}, \delta^{\perp} \rangle \leq \|\delta^{\perp}\|^2$. Using (8b) implies that $\|\Lambda^{-1}\|_2 \leq 1$ thus we obtain $\|p_{\Lambda}^{\perp}\| \leq \|p^{\perp}\| \leq \|\delta^{\perp}\|$ and $\langle \delta^{\perp}, p_{\Lambda}^{\perp} \rangle \leq \|\delta^{\perp}\|^2$. Therefore, $\tfrac{1}{4}\|\delta^{\perp}\|^2 + \tfrac{1}{2}\langle \delta^{\perp}, p_{\Lambda}^{\perp} \rangle + \tfrac{1}{4}\|p_{\Lambda}^{\perp}\|^2 \leq \tfrac{1}{4}\|\delta^{\perp}\|^2 + \tfrac{1}{2}\|\delta^{\perp}\|^2 + \tfrac{1}{4}\|\delta^{\perp}\|^2 = \|\delta^{\perp}\|^2$, i.e., the result holds for the range space portion.

ii) We consider all δ^0, p^0 such that $0 \leq \|p^0\|^2 \leq \langle p^0, \delta^0 \rangle \leq \|\delta^0\|^2$. It follows that $\|\delta^0\|^2 - 2\langle \delta^0, p^0 \rangle + \|p^0\|^2 \leq \|\delta^0\|^2 - 2\langle \delta^0, p^0 \rangle + \langle \delta^0, p^0 \rangle \leq \|\delta^0\|^2$. Thus, the result also holds for the null space portion.
Cases i) and ii) together imply the result. $\qquad \qquad \square$

Appendix B. Condition for satisfying (A3). We give an algorithm to check whether Assumption (A3) is satisfied at every point in $\mathcal{Z}$, representing a necessary and sufficient condition. This conditions relies on the enumeration of all possible combinations of active components and their active sets, and can be checked once and "offline" for all parameters θ. We consider $\mathcal{E} := \{z \in \mathbb{R}^n \mid Az = b\}$, with $A \in \mathbb{R}^{m \times n}$, and $\mathcal{Z}(\theta)$ with components $\mathcal{Z}^i(\theta)$ defined as

$$(23) \qquad \mathcal{Z}(\theta) = \bigcup_{i=1}^{I} \mathcal{Z}^i(\theta) = \bigcup_{i=1}^{I} \{z \in \mathbb{R}^n \mid G^i z = g^i(\theta), F^i z \leq f^i(\theta)\} ,$$

where $G^i \in \mathbb{R}^{p_i \times n}$ and $F^i \in \mathbb{R}^{m_i \times n}$. The functions $f^i(\theta) = F^{i,\theta}\theta + f^i$ and $g^i(\theta) = G^{i,\theta}\theta + g^i$ depend affinely on a parameter $\theta \in \mathbb{R}^p$. With $\mathcal{Z}^i(\theta)$ closed convex polyhedra for fixed $\theta \in \mathbb{R}^p$ and non-empty for some θ.

Given a parameter θ and index $i \in \{1, \ldots, I\}$. The *active set* $\mathcal{J}_{\mathcal{Z}^i(\theta)}(z) := \{j \in \{1, \ldots, m_i\} \mid$ s.t. $F_j^i z = f_j^i(\theta)\}$ of $\mathcal{Z}^i(\theta)$ at $z \in \mathcal{Z}^i(\theta)$ is defined in the usual way [27, Definition 12.1, p. 308], where F_j^i denotes the j-th row of F^i. For $z \notin \mathcal{Z}^i(\theta)$, we define $\mathcal{J}_{\mathcal{Z}^i(\theta)}(z) := \varnothing$. For any active set $\mathcal{J} \subseteq \{1, \ldots, m_i\}$, $F_{\mathcal{J}}^i$ denotes the subset of rows of F^i that belong to the active set $\mathcal{J}$, $f_{\mathcal{J}}^i$ is defined in the same way. We further define the active set $\mathcal{J}_{\mathcal{Z}(\theta)}(z)$ of $\mathcal{Z}(\theta)$ at $z \in \mathcal{Z}(\theta)$, as the collection of all the active sets of the components of $\mathcal{Z}(\theta)$, i.e., $\mathcal{J}_{\mathcal{Z}(\theta)}(z) := \{\mathcal{J}_{\mathcal{Z}^i(\theta)}(z)\}_{i=1}^I$,

The set $\mathcal{I}_{\mathcal{Z}} \subseteq 2^{\{1, \ldots, I\}}$ denotes the set of sets of active components of $\mathcal{Z}(\theta)$, i.e., $\mathcal{I} \in \mathcal{I}_{\mathcal{Z}}$ if and only if there exists $\theta \in \mathbb{R}^p$ and $z \in \mathcal{Z}(\theta)$ such that $\mathcal{I} \equiv \mathcal{I}_{\mathcal{Z}(\theta)}(z)$, where $\mathcal{I}_{\mathcal{Z}(\theta)}(z)$ denotes the set of active components at z, as defined in Section 4.2. Similarly we denote by $\mathcal{J}_{\mathcal{Z}} := \{\mathcal{J}_{\mathcal{Z}^i}\}_{i=1}^I$ the collection of the sets of all possible active sets of $\mathcal{Z}(\theta)$, where $\mathcal{J}_{\mathcal{Z}^i} := \{\mathcal{J} \subseteq \{1, \ldots, m_i\} \mid \exists \theta \in \mathbb{R}^p, \exists z \in \mathcal{Z}(\theta) \text{ s.t. } \mathcal{J} \equiv \mathcal{J}_{\mathcal{Z}^i(\theta)}(z)\}$ denotes the set of all possible active sets for component i.

Assumption (A3) is purely geometric and needs to be satisfied at every point in $\mathcal{Z}$. Algorithm 3 will effectively enumerate all possible combinations of active components and their active sets and check for each such combination. This amounts to a finite number of checks. Checking a parametric set $\mathcal{Z}(\theta)$, due to the polyhedral structure, requires checking finitely many different non-parametric sets.

Algorithm 3 Check of Assumption (A3)

Require: $\mathcal{Z}(\theta)$ as in (23)
 1: **for each** $\mathcal{I} \in \mathcal{I}_{\mathcal{Z}}, \{\mathcal{J}_i\}_{i=1}^I \in \mathcal{J}_{\mathcal{Z}}$ **do**
 2:     $\mathcal{N} \leftarrow \bigcap_{i \in \mathcal{I}} \{v \mid v = G^{i\top}\nu + F_{\mathcal{J}_i}^{i\top}\mu, \mu \geq 0\}$ ▷ construct regular normal cone
 3:     **if** $\mathcal{N} = \{0\}$ **then continue**
 4:     $\mathcal{R} \leftarrow \bigcup_{i \in \mathcal{I}} \{v \mid G^i v = 0, F_{\mathcal{J}_i}^i v \leq 0\}$ ▷ construct union of recession cones
 5:     **if** $\mathcal{N}^{\perp} \subseteq \mathcal{R}$ **then continue**
 6:     **else** $w \leftarrow w \in \mathcal{N}^{\perp} \setminus \mathcal{R}$ **return** $(w, \{\mathcal{A}_j\}_{j=1}^m)$ ▷ return counter-example

---

Algorithm 3 is a combinatorial algorithm and may perform badly for anything but small dimensions. However when the set $\mathcal{Z}$ is a Cartesian product of sets in low dimension, the check can be performed separately for each part of the Cartesian product. This significantly reduces the "curse of dimensionality".

LEMMA B.1. *Given a (parametric) set $\mathcal{Z}(\theta)$ as in (23), then Algorithm 3 terminates in a finite number of steps. Either $\mathcal{Z}(\theta)$ satisfies Assumption (A3) for all θ, or the algorithm returns a valid counter-example.*

Proof of Lemma B.1. We will show that Algorithm 3 effectively checks Assumption (A3) for every θ and every point in $\mathcal{Z}(\theta)$. We will argue, that checking (A3) for every $\theta \in \mathbb{R}^p$ and every $z \in \mathcal{Z}(\theta)$ is equivalent to checking it for every set of active constraints. Note that $\mathcal{J}_{\mathcal{Z}}$ the set of active sets of $\mathcal{Z}(\theta)$ is a finite set with at most $2^{\sum_{i=1}^I m_i}$ elements.

In Algorithm 3, we iterate over the elements of $\mathcal{I}_{\mathcal{Z}}$ and $\mathcal{J}_{\mathcal{Z}}$, which means that the loop is executed at most a finite number of times. Given a pair $\mathcal{I}, \{\mathcal{J}_i\}_{i=1}^I$ we want to verify, or invalidate, (A3) by considering two sets $\mathcal{N}$ and $\mathcal{R}$, these set are defined solely via the active constraints and do not depend on θ. For each active component

$i \in \mathcal{I}$ we define two auxiliary sets

$$\mathcal{A}_i(\theta) := \{z \mid G^i z = g^i(\theta), F^i_{\mathcal{J}_i} z = f^i_{\mathcal{J}_i}(\theta), F^i_{\mathcal{J}_i^c} z < f^i_{\mathcal{J}_i^c}(\theta)\},$$
$$\mathcal{F}_i(\theta) := \{z \mid G^i z = g^i(\theta), F^i_{\mathcal{J}_i} z \le f^i_{\mathcal{J}_i}(\theta)\},$$

where the set $\mathcal{J}_i^c$ is simply the complement of $\mathcal{J}_i$, i.e., the set of constraints that are not active. For a given θ, $\mathcal{A}_i(\theta)$ is the set of points $z \in \mathcal{Z}$ that belong to the active set $\mathcal{J}_i$, whereas $\mathcal{F}_i(\theta) \supseteq \mathcal{A}_i(\theta)$ is a larger set, that is defined by only the active constraints and in general is not included in $\mathcal{Z}$. We further define $\mathcal{A}(\theta) := \bigcap_{i=1}^I \mathcal{A}_i(\theta)$, the set of points that belong to all of the active sets of the active components. By checking Assumption (A3) for each $\mathcal{A}(\theta)$ we will effectively check it for each $z \in \mathcal{Z}$.

The conditions to be checked involve the regular normal cone $\hat{\mathcal{N}}_{\mathcal{Z}(\theta)}(z)$. For each $\theta, z \in \mathcal{A}(\theta)$ the regular normal cone $\hat{\mathcal{N}}_{\mathcal{Z}(\theta)}(z)$ is given as $\hat{\mathcal{N}}_{\mathcal{Z}(\theta)}(z) = \bigcap_{i \in \mathcal{I}} \{G^{i\top} \nu + F^{i\top}_{\mathcal{J}_i} \mu \mid \mu \ge 0\}$, using Proposition A.1 and [29, Theorem 6.46, p. 231]. It is independent of θ and z, and only depends the active constraints and components. This set, $\mathcal{N} := \hat{\mathcal{N}}_{\mathcal{Z}(\theta)}(z)$, is computed in step 2 of Algorithm 3. A half-space representation of $\mathcal{N}$ can be obtained using Fourier-Motzkin elimination [8, p. 84] for each $i \in \mathcal{I}$, after which the intersections are trivial. We now need to check two conditions. If $\mathcal{N} = \{0\}$, then the first part of (A3) is satisfied for all $z \in \mathcal{A}(\theta)$, and we continue with the next active set. Otherwise, we need to check whether there exists an $\epsilon > 0$ such that for all $w \in \mathcal{N}^\perp \cap \mathcal{B}_\epsilon$ we have that $z + w \in \mathcal{Z}(\theta)$. In order to check this condition, in step 4, we construct the set $\mathcal{R}$ as the union of the recession cones $\mathcal{R}_i := \operatorname{rec}(\mathcal{F}_i(\theta)) = \{v \mid G^i v = 0, F^i_{\mathcal{J}_i} v \le 0\}$ of $\mathcal{F}_i(\theta)$. Finally it remains to show that checking this condition for all $z \in \mathcal{A}(\theta)$, in step 5, is equivalent to checking whether the orthogonal complement $\mathcal{N}^\perp := \{v \in \mathbb{R}^n \mid \langle v, u \rangle = 0, \forall u \in \mathcal{N}\}$ is contained in $\mathcal{R}$. If $\mathcal{N}^\perp \subseteq \mathcal{R}$, consider any $z \in \mathcal{A}(\theta)$, an $\epsilon > 0$ small enough and any $w \in \mathcal{N}^\perp \cap \mathcal{B}_\epsilon$. Clearly there exists an $i \in \mathcal{I}$ such that $w \in \mathcal{R}_i$. Furthermore, by construction of $\mathcal{A}(\theta)$ and $\mathcal{R}_i$, we have $z + w \in \mathcal{Z}^i \subseteq \mathcal{Z}$. This means (A3) is satisfied for all $z \in \mathcal{A}(\theta)$ and we continue with the next active set. If $\mathcal{N}^\perp \not\subseteq \mathcal{R}$, then there exists a $w \in \mathcal{N}^\perp \setminus \mathcal{R}$. Because $\mathcal{N}$ and $\mathcal{R}_i$ are cones, this is true for all positive scalings of w. In particular, for any $\epsilon > 0$ there exists a $t > 0$ such that $tw \in \mathcal{N}^\perp \cap \mathcal{B}_\epsilon$ and $tw \notin \mathcal{R}$. Analogously to above, it follows that for any $z \in \mathcal{A}(\theta)$, $z + tw \notin \mathcal{Z}^i$ for all $i \in \mathcal{I}$, and therefore $z + tw \notin \mathcal{Z}$, which violates (A3). In this case, the algorithm terminates and returns a violating direction w and the corresponding active set, from which violating z, θ can be reconstructed. Because the Algorithm 3 enumerates all active sets, it either terminates with a counterexample, or it returns after finitely many steps, having checked all possibilities. $\qquad\square$

REFERENCES

[1] A. ALESSIO AND A. BEMPORAD, *A Survey on Explicit Model Predictive Control*, Springer Berlin Heidelberg, 2009, pp. 345–369.

[2] A. ARCE, A. DEL REAL, AND C. BORDONS, *MPC for battery/fuel cell hybrid vehicles including fuel cell dynamics and battery performance improvement*, Journal of Process Control, 19 (2009), pp. 1289–1304.

[3] D. AXEHILL, *Integer Quadratic Programming for Control and Communication*, PhD thesis, Linköping University, 2008.

[4] H. BAUSCHKE AND P. COMBETTES, *Convex analysis and monotone operator theory in Hilbert spaces*, Springer, 2011.

[5] A. BEMPORAD AND M. MORARI, *Control of systems integrating logic, dynamics, and constraints*, Automatica, 35 (1999), pp. 407–427.

[6] V. BERINDE, *Iterative approximation of fixed points*, Springer, 2 ed., 2007.

[7] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Foundations and Trends in Machine Learning, 3 (2011), pp. 1–122.

[8] G. Dantzig, *Linear programming and extensions*, Princeton University Press, 1963.

[9] S. Di Cairano, A. Bemporad, I. Kolmanovsky, and D. Hrovat, *Model predictive control of magnetically actuated mass spring dampers for automotive applications*, International Journal of Control, 80 (2007), pp. 1701–1716.

[10] A. Domahidi, A. Zgraggen, M. Zeilinger, M. Morari, and C. Jones, *Efficient interior point methods for multistage problems arising in receding horizon control*, in IEEE Conference on Decision and Control, 2012, pp. 668–674.

[11] D. Frick, *Automatic code-generation tool.* http://github.com/dafrick/codegen, 2016.

[12] D. Frick, A. Domahidi, and M. Morari, *Embedded optimization for mixed logical dynamical systems*, Computers & Chemical Engineering, 72 (2015), pp. 21–33.

[13] T. Geyer, G. Papafotiou, R. Frasca, and M. Morari, *Constrained optimal control of the step-down dc-dc converter*, IEEE Transactions on Power Electronics, 23 (2008), pp. 2454–2464.

[14] A. Hempel, *Control of Piecewise Affine Systems Through Inverse Optimization*, PhD thesis, ETH Zürich, 2015. Dissertation, ETH-Zürich, 2015, No. 23222.

[15] R. Henrion and J. Outrata, *On calculating the normal cone to a finite union of convex polyhedra*, Optimization, 57 (2008), pp. 57–78.

[16] M. Herceg, M. Kvasnica, C. Jones, and M. Morari, *Multi-parametric toolbox 3.0*, in European Control Conference, 2013, pp. 502–510.

[17] J.-H. Hours and C. Jones, *A parametric nonconvex decomposition algorithm for real-time and distributed NMPC*, IEEE Transactions on Automatic Control, 61 (2016), pp. 287–302.

[18] B. Houska, J. Frasch, and M. Diehl, *An augmented lagrangian based algorithm for distributed nonconvex optimization*, SIAM Journal on Optimization, 26 (2016), pp. 1101–1127.

[19] IBM, *IBM ILOG CPLEX optimization studio*, 2015.

[20] M. Jung, *Relaxations and Approximations for Mixed-Integer Optimal Control*, PhD thesis, Universität Heidelberg, 2013.

[21] R. Lang, *A note on the measurability of convex sets*, Archiv der Mathematik, 47 (1986), pp. 90–92.

[22] G. Li and T. Pong, *Global convergence of splitting methods for nonconvex composite optimization*, SIAM Journal on Optimization, 25 (2015), pp. 2434–2460.

[23] A. Liniger, A. Domahidi, and M. Morari, *Optimization-based autonomous racing of 1:43 scale RC cars*, Optimal Control Applications and Methods, 36 (2015), pp. 628–647.

[24] J. Löfberg, *YALMIP : a toolbox for modeling and optimization in MATLAB*, in IEEE International Symposium on Computer Aided Control Systems Design, 2004, pp. 284–289.

[25] S. Magnússon, P. C. Weeraddana, M. G. Rabbat, and C. Fischione, *On the convergence of alternating direction lagrangian methods for nonconvex structured optimization problems*, IEEE Transactions on Control of Network Systems, 3 (2016), pp. 296–309.

[26] M. Morari and J. Lee, *Model predictive control: past, present and future*, Computers & Chemical Engineering, 23 (1999), pp. 667–682.

[27] J. Nocedal and S. Wright, *Numerical optimization*, Springer Science+Business Media, 2 ed., 2006.

[28] R. A. Poliquin, R. T. Rockafellar, and L. Thibault, *Local differentiability of distance functions*, Transactions of the American Mathematical Society, 352 (2000), pp. 5231–5249.

[29] R. Rockafellar and R. Wets, *Variational analysis*, Springer, 1998.

[30] A. Ruszczyński, *Nonlinear Optimization*, Princeton University Press, 2006.

[31] S. Sager, *Numerical methods for mixed-integer optimal control problems*, PhD thesis, Universität Heidelberg, 2005.

[32] R. Takapoui, N. Moehle, S. Boyd, and A. Bemporad, *A simple effective heuristic for embedded mixed-integer quadratic programming*, International Journal of Control, 0 (2017), pp. 1–11.

[33] J. Vielma, *Mixed integer linear programming formulation techniques*, SIAM Review, 57 (2015), pp. 3–57.

[34] J. Ye, *Constraint qualifications and necessary optimality conditions for optimization problems with variational inequality constraints*, SIAM Journal on Optimization, 10 (2000), pp. 943–962.