Official URL

DOI : https://doi.org/10.1137/16M1133889

SIAM J. OPTIM.
Vol. 28, No. 2, pp. 1867–1891

QUALIFICATION CONDITIONS IN SEMIALGEBRAIC PROGRAMMING*

JÉRÔME BOLTE†, ANTOINE HOCHART‡, AND EDOUARD PAUWELS§

Abstract. For an arbitrary finite family of semialgebraic/definable functions, we consider the corresponding inequality constraint set and we study qualification conditions for perturbations of this set. In particular we prove that all positive diagonal perturbations, save perhaps a finite number of them, ensure that any point within the feasible set satisfies the Mangasarian–Fromovitz constraint qualification. Using the Milnor–Thom theorem, we provide a bound for the number of singular perturbations when the constraints are polynomial functions. Examples show that the order of magnitude of our exponential bound is relevant. Our perturbation approach provides a simple protocol to build sequences of "regular" problems approximating an arbitrary semialgebraic/definable problem. Applications to sequential quadratic programming methods and sum of squares relaxation are provided.

Key words. constraint qualification, Mangasarian–Fromovitz, Arrow–Hurwicz–Uzawa, Lagrange multipliers, optimality conditions, tame programming

AMS subject classifications. Primary, 26D10; Secondary, 32B20, 49K24, 49J52, 37B35, 14P15

DOI. 10.1137/16M1133889

1. Introduction.

Constraint qualification conditions ensure that normal cones are finitely generated by the gradients of the active constraints. When considering an optimization problem, this fact immediately provides Lagrange/KKT necessary optimality conditions, which are at the root of most resolution methods (see, e.g., [6, 42]). Finding settings in which qualification conditions are easy to formulate and easy to verify is thus of fundamental importance. In a convex framework, the power of Slater's condition consists in its extreme simplicity: the resolution of a "simple" problem (e.g., finding an interior point), often done directly or through routine computations, guarantees the regularity of the problem.

In a nonconvex setting, the question becomes much more delicate but the wish is the same: to describe normal cones as gradient-generated cones in order to derive KKT conditions (see, e.g., [43]). Contrary to what happens for convex functions, the knowledge of the functions at one point does not capture enough information about the global geometry to infer well-posedness everywhere.[1] Very smooth and simple problems satisfying all possible natural conditions can generally present a failure of qualification, for which the normal cone is not generated by the gradients of the active

*Received by the editors June 9, 2017; accepted for publication (in revised form) February 26, 2018; published electronically June 28, 2018.
http://www.siam.org/journals/siopt/28-2/M113388.html
Funding: This work was sponsored by the Air Force Office of Scientific Research, Air Force Material Command, USAF, under grant number FA9550-15-1-0500. This work was also partially supported by PGMO.
†Toulouse School of Economics, Université Toulouse 1 Capitole, Toulouse, France (jerome.bolte@tse-fr.eu).
‡Universidad Adolfo Ibáñez, Santiago, Chile (antoine.hochart@gmail.com).
§Institut de Recherche en Informatique de Toulouse, Université Paul Sabatier, Toulouse, France (edouard.pauwels@irit.fr).

[1] Observe though that the local knowledge of a polynomial function implies the knowledge of the function everywhere. But, as far as we know, this fact has never given birth to any simple qualification condition.

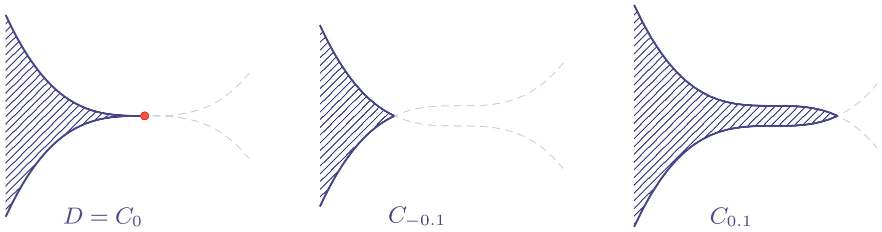

FIG. 1. *On the left, the constraint set D (see* (1.1)*): the bullet highlights a cusp, for us a failure of constraint qualification. Middle and right: negative and positive perturbations of $D = C_0$ make the cusp disappear.*

constraints, and thus KKT conditions cannot apply. In dimension 2 a typical failure is a cusp, illustrated in Figure 1 for the constraint set

$$D = \left\{(x_1, x_2) \in \mathbb{R}^2 \mid x_1^3 + x_2 \leqslant 0,\ x_1^3 - x_2 \leqslant 0\right\}. \tag{1.1}$$

Since in this general setting simple qualification conditions are not available, several researchers have considered the problem under the angle of perturbations. To our knowledge, the first work in this direction was proposed by Spingarn and Rockafellar [47]. Given differentiable functions $g_1, \ldots, g_m : \mathbb{R}^n \to \mathbb{R}$ and a constraint set $C = \{x \in \mathbb{R}^n \mid g_1(x) \leqslant 0, \ldots, g_m(x) \leqslant 0\}$, they indeed introduced the perturbed constraint sets

$$C_\mu := [g_1 \leqslant \mu_1, \ldots, g_m \leqslant \mu_m], \quad \text{where } \mu = (\mu_1, \ldots, \mu_m) \in \mathbb{R}^m,$$

and studied their properties regarding qualification conditions. In what follows, a set C_μ for which qualification conditions hold at each feasible point is said to be *regular*. Accordingly the corresponding perturbation μ is called *regular*. When $m = 1$, one obviously recovers the usual definition of a regular value of a function (see, e.g., Milnor's monograph [38]), and one guesses that a major role will be played by Sard-type theorems. Recall that Sard's original theorem (see, e.g., [38]) expresses that the regular values of a sufficiently smooth function are generic within $\mathbb{R}^m$. For $m \geqslant 1$, the work on perturbed constraint sets by Spingarn and Rockafellar [47] dealt with the genericity of regular values using a quite restrictive notion of qualification condition. Works by Fujiwara [29], Scholtes and Stöhr [45], and Nie [41] gave further insights on other different aspects but with the same type of qualification assumptions. When the mappings g_i are semialgebraic (or definable), the application of the definable nonsmooth Sard's theorem of Ioffe [32] yields stronger results since, in that case, regularity exactly corresponds to sets satisfying the Mangasarian–Fromovitz constraint qualification everywhere (see Theorem 2.9). These aspects are discussed in detail in section 2.

Genericity results have been the object of a recent revival in connection with semialgebraic optimization: see Bolte, Daniilidis, and Lewis [11], Daniilidis and Pang [18], Drusvyatskiy, Ioffe, and Lewis [25], Lee and Pham [35], and Hà and Pham [31]. An original feature of our work is to exploit the fact that genericity is a relative concept. A property is indeed generic within some given family, but if one considers smaller families, genericity may no longer hold. It is therefore important to identify the smallest possible families in order to strengthen genericity results and to be able to exploit them for improving effective optimization techniques (e.g., algorithms, homotopy methods). In this regard, we address in section 3 the following two questions.

– *How do we perturb to ensure regularity? In other words, how can we build simple problems* $(\mathcal{P}_\alpha)_{\alpha\in\mathbb{R}_+}$ *which are regular and whose values,* $\mathrm{val}(\mathcal{P}_\alpha)$*, converge to that of the original problem,* $\mathrm{val}(\mathcal{P}_0)$*?*
– *Can we go beyond mere genericity and quantify the number of singular (i.e., nonregular) values in the polynomial case?*

Our first result, à la Morse–Sard, relies on definability assumptions of the data (e.g., semialgebraicity) and provides one-parameter families of regular constraint sets. This is done by showing that any positive semiline $\mathbb{R}_+ v$, with $v \in \mathbb{R}^m_{++}$, bears only finitely many singular perturbations. For instance if we let $\boldsymbol{\alpha} := (\alpha, \ldots, \alpha)$, the sets $(C_{\boldsymbol{\alpha}})_{\alpha\in\mathbb{R}_+}$ are regular for all α positive small enough; see Figure 1 for an illustration. When some of the constraint functions are convex, our approach is considerably simplified: we show indeed that a "partial Slater condition" allows us to restrict the perturbation approach to nonconvex functions.

The strength of our results is well conveyed by the following general approximation fact: for any objective f and for α small enough, we are able to build explicit well-posed problems

$$(\mathcal{P}_\alpha) \qquad \text{minimize } f(x) \quad \text{subject to (s.t.) } x \in C_{\boldsymbol{\alpha}}$$

which satisfy, under mild conditions, $\lim_{\alpha\to 0^+} \mathrm{val}(\mathcal{P}_\alpha) = \mathrm{val}(\mathcal{P}_0)$. Our approach opens the way to continuation methods (see [2] and references therein) or to more direct diagonal methods as shown in our final section.

A natural question which immediately emerges is whether it is possible to count the number of singular perturbations. When assuming further that the data are polynomial functions whose degrees are bounded by d, we show by using the Milnor–Thom theorem that the number of singular values for problems of the type $(\mathcal{P}_\alpha)$ is lower than $d(2d-1)^n(2d+1)^m$. Examples show that in general the bound is indeed exponential, even in the quadratic case with only one of the g_i being nonconvex. The worst-case bound described in this work is a rather negative result for semialgebraic programming in the sense that it shows that there are instances for which singular values are so clumped and numerous that perturbation techniques are ineffective. That worst-case instances of general semialgebraic programming are out of reach of modern methods became a well-known fact due to the pioneering work [7]. It would be interesting to recast our findings from this perspective. For instance, our results suggest that some constraint sets might have such a complex nature that most local methods are inapplicable in practice, even after perturbation. On the other hand, as suggested by real-life problems, regular instances are numerous in practice. This shows the need to understand further the geometric factors or probabilistic priors on the constraints that could make singular values less numerous or at least favorably distributed.

In section 4, we provide two theoretical algorithmic illustrations of our results. As a general fact, our diagonal perturbation scheme can be used in conjunction with any algorithm whose behavior relies on constraint qualification assumptions. We illustrate this principle with exact semidefinite programming relaxations in polynomial programming, for which well-behaved constructions were proposed for regular problems [1, 19, 20]. A second application of our general results is given by a class of sequential quadratic programming methods, SQP for short. SQP methods are widespread in practical applications; see, e.g., [3, 13, 26]. Convergence analysis of such methods usually requires very strong qualification conditions in order to handle regularity and infeasibility issues for minimizing sequences. We show how our perturbation results provide a natural and strong tool for convergence analysis in the framework of semialgebraic optimization.

2. Regular and singular perturbations of constraint sets.

2.1. Notation and definitions.

Constraint sets and qualification conditions. Let us consider the general nonlinear optimization problem

$$(\mathcal{P}_{\text{NLP}}) \qquad \begin{aligned} \text{minimize } & f(x) \\ \text{subject to } & g_1(x) \leqslant 0, \dots, g_m(x) \leqslant 0, \\ & h_1(x) = 0, \dots, h_r(x) = 0, \end{aligned}$$

where f, $g_1, \dots, g_m, h_1, \dots, h_r$ are differentiable functions from $\mathbb{R}^n$ to $\mathbb{R}$. We denote by

$$C \;=\; [g_1 \leqslant 0, \dots, g_m \leqslant 0] \;:=\; \{x \in \mathbb{R}^n \mid g_1(x) \leqslant 0, \dots, g_m(x) \leqslant 0\}$$

the inequality constraint set and by

$$M \;=\; [h_1 = 0, \dots, h_r = 0] \;:=\; \{x \in \mathbb{R}^n \mid h_1(x) = 0, \dots, h_r(x) = 0\}$$

the equality constraint set. For $x \in C$, we define the set of active constraints by

$$I(x) \;:=\; \{1 \leqslant i \leqslant m \mid g_i(x) = 0\}.$$

We next recall a standard regularity condition.

DEFINITION 2.1 (Mangasarian–Fromovitz constraint qualification). *A point $x \in C \cap M$ is said to satisfy the* Mangasarian–Fromovitz constraint qualification (MFCQ) *if the gradient vectors $\nabla h_j(x)$, $j = 1, \dots, r$, are linearly independent and there exists $y \in \mathbb{R}^n$ such that*

$$\begin{cases} \langle y, \nabla h_j(x) \rangle = 0, & j = 1, \dots, r, \\ \langle y, \nabla g_i(x) \rangle < 0, & i \in I(x). \end{cases} \tag{2.1}$$

If there is no equality constraint, this condition is then called the Arrow–Hurwicz–Uzawa constraint qualification. *We say that MFCQ holds throughout $C \cap M$ if it is satisfied at every point in $C \cap M$.*

Remark 2.2. By a straightforward application of the Hahn–Banach separation theorem, the existence of a vector $y \in \mathbb{R}^n$ satisfying condition (2.1) is equivalent to

$$\operatorname{co}\{\nabla g_i(x) \mid i \in I(x)\} \;\cap\; \operatorname{span}\{\nabla h_j(x) \mid 1 \leqslant j \leqslant r\} = \emptyset,$$

where $\operatorname{co} X$ denotes the convex hull of any subset $X \subset \mathbb{R}^n$, and $\operatorname{span} X$ its linear span. If there is no equality constraint, this characterization simply reads

$$0 \notin \operatorname{co}\{\nabla g_i(x) \mid i \in I(x)\}.$$

Let us briefly recall that MFCQ guarantees the existence of Lagrange multipliers at minimizers of problem $(\mathcal{P}_{\text{NLP}})$: if a local minimizer $\bar{x}$ of f on $C \cap M$ satisfies MFCQ, then there exist multipliers $\lambda_1, \ldots, \lambda_m \in \mathbb{R}_+ := [0, +\infty)$ and $\kappa_1, \ldots, \kappa_r \in \mathbb{R}$ such that

$$\begin{cases} \nabla f(\bar{x}) + \displaystyle\sum_{i=1}^{m} \lambda_i \nabla g_i(\bar{x}) + \sum_{j=1}^{r} \kappa_j \nabla h_j(\bar{x}) = 0, \\ \lambda_i\, g_i(\bar{x}) = 0, \quad i = 1, \ldots, m. \end{cases} \tag{2.2}$$

Any feasible point satisfying these conditions is called a *Karush–Kuhn–Tucker (KKT) point*.

Remark 2.3 (Clarke regularity and MFCQ). A more geometrical way of formulating the existence of Lagrange multipliers consists in interpreting the gradients of active constraints as generators of a cone normal to the constraint set. In the terminology of modern nonsmooth analysis, assuming that there are only inequality constraints in problem $(\mathcal{P}_{\text{NLP}})$, this amounts to the normal regularity of the set $C = [g_1 \leqslant 0, \ldots, g_m \leqslant 0]$. We next explain this fact.

Given a nonempty closed subset $X \subset \mathbb{R}^n$, the *Fréchet normal cone* to X at point $\bar{x} \in X$ is defined by

$$\hat{N}_X(\bar{x}) := \{v \in \mathbb{R}^n \mid \langle v, x - \bar{x}\rangle \leqslant o(\|x - \bar{x}\|),\ x \in X\}.$$

It is immediate to prove that any solution to problem $(\mathcal{P}_{\text{NLP}})$ satisfies

$$\nabla f(x) + \hat{N}_C(x) \ni 0, \tag{2.3}$$

which suggests expressing $\hat{N}_C(x)$ in terms of the initial data $g_1, \ldots, g_m$. To do so, let us introduce the *limiting normal* or *Mordukhovich normal cone*[2] to X at $\bar{x}$, denoted by $N_X(\bar{x})$ and defined by

$$v \in N_X(\bar{x}) \iff \exists\, x_n \to \bar{x}, \quad \exists\, v_n \to v, \quad v_n \in \hat{N}_X(x_n).$$

The set X is called *regular* at $\bar{x}$ if $\hat{N}_X(\bar{x}) = N_X(\bar{x})$.

By classical results of nonsmooth analysis, if the Mangasarian–Fromovitz constraint qualification holds throughout C, then C is regular at every point in C (see [43, Thm. 6.14] or [15, Thm. 7.2.6]). In addition, we have, for all $x \in C$,

$$N_C(x) = \left\{ \sum_{i \in I(x)} \lambda_i\ \nabla g_i(x) \;\middle|\; \lambda_i \geqslant 0,\ i \in I(x) \right\},$$

which combined with (2.3) yields the claimed result.

Perturbations of constraint sets. For $\mu \in \mathbb{R}^m$ and $\nu \in \mathbb{R}^r$, we denote by

$$\begin{aligned} C_\mu &:= [g_1 \leqslant \mu_1, \ldots, g_m \leqslant \mu_m] = \{x \in \mathbb{R}^n \mid g_1(x) \leqslant \mu_1, \ldots, g_m(x) \leqslant \mu_m\}, \\ M_\nu &:= [h_1 = \nu_1, \ldots, h_r = \nu_r] = \{x \in \mathbb{R}^n \mid h_1(x) = \nu_1, \ldots, h_r(x) = \nu_r\} \end{aligned}$$

the perturbed inequality and equality constraint sets of problem $(\mathcal{P}_{\text{NLP}})$, respectively. Also, we denote by

$$\mathcal{A} := \{(\mu, \nu) \in \mathbb{R}^m \times \mathbb{R}^r \mid C_\mu \cap M_\nu \neq \emptyset\}$$

the set of *admissible perturbations*.

[2] See the pioneering work [39].

Definition 2.4 (regular/singular perturbations). *We say that* $(\mu,\nu) \in \mathcal{A}$ *is a* regular perturbation *if MFCQ holds throughout* $C_\mu \cap M_\nu$ *and we denote by* $\mathcal{A}_{\mathrm{reg}}$ *the collection of all regular perturbations:*

$$\mathcal{A}_{\mathrm{reg}} := \{(\mu,\nu) \in \mathcal{A} \mid \textit{MFCQ holds at every } x \in C_\mu \cap M_\nu\}.$$

In contrast, an admissible perturbation $(\mu,\nu) \in \mathcal{A}$ *is* singular *if MFCQ is not satisfied at some point of* $C_\mu \cap M_\nu$. *The subset of singular perturbations is given by*

$$\mathcal{A}_{\mathrm{sing}} := \mathcal{A} \setminus \mathcal{A}_{\mathrm{reg}}.$$

Up to an obvious change of definition, we shall use the same notation when there is no equality constraint.

2.2. Metric regularity and constraint qualification. In this subsection, we recall how the Mangasarian–Fromovitz constraint qualification can be interpreted in terms of metric regularity of some set-valued mapping. For that purpose, we gather below some classical notions in nonsmooth analysis (see [22, 40, 43]).

A set-valued mapping $F : \mathbb{R}^p \rightrightarrows \mathbb{R}^q$ is a map sending each point of $\mathbb{R}^p$ to a subset of $\mathbb{R}^q$. We denote by $\operatorname{graph} F := \{(x,y) \in \mathbb{R}^p \times \mathbb{R}^q \mid y \in F(x)\}$ the *graph* of F and by $\operatorname{dom} F := \{x \in \mathbb{R}^p \mid F(x) \neq \emptyset\}$ its *domain*.

The set-valued mapping $F : \mathbb{R}^p \rightrightarrows \mathbb{R}^q$ is *metrically regular* at $(\bar{x},\bar{y}) \in \operatorname{graph} F$ if the graph of F is locally closed at $(\bar{x},\bar{y})$ and there exists a positive real number κ, together with neighborhoods $\mathcal{U}$ and $\mathcal{V}$ of $\bar{x}$ and $\bar{y}$, respectively, such that

$$\operatorname{dist}(x, F^{-1}(y)) \leqslant \kappa \operatorname{dist}(y, F(x))$$

for all $(x,y) \in \mathcal{U} \times \mathcal{V}$. Here, $\operatorname{dist}(z,K)$ refers to the distance of any point z of a space endowed with a norm $\|\cdot\|$ to any subset K of the same space, i.e., $\inf_{k\in K} \|k - z\|$.

We now come back to problem $(\mathcal{P}_{\mathrm{NLP}})$ and introduce the set-valued mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^{m+r}$ defined by

$$F(x) = \begin{pmatrix} g_1(x) \\ \vdots \\ g_m(x) \\ h_1(x) \\ \vdots \\ h_r(x) \end{pmatrix} + \mathbb{R}^m_+ \times \{0\}^r, \tag{2.4}$$

where $\mathbb{R}^m_+$ is the nonnegative orthant of $\mathbb{R}^m$. Observe that $(\mu,\nu) \in F(x)$ if and only if $x \in C_\mu \cap M_\nu$. Also notice that, by continuity of the constraint functions, $\operatorname{graph} F$ is closed.

The following result, due to Robinson, characterizes the points satisfying MFCQ in terms of the mapping F. For a thorough discussion and various proofs, we refer the reader to [40]. Other approaches are [43, Ex. 9.44], [22, Ex. 4F.3], or [21, Thm. 4.1], which avoid the use of coderivative calculus.

Theorem 2.5 (Robinson). *The Mangasarian–Fromovitz constraint qualification holds at point* $x \in C_\mu \cap M_\nu$ *if and only if the set-valued mapping* F *defined in* (2.4) *is metrically regular at* $(x,(\mu,\nu))$.

2.3. Genericity of regular perturbations. Qualification conditions play an important role in the analysis of nonlinear programming and the convergence of optimization algorithms, yet checking these conditions at optimal points is hardly possible. This is why one instead seeks local/global simple geometrical assumptions that automatically warrant these conditions. Sard's theorem provides results in this direction: generic equations are well posed if the data are smooth enough or well structured (e.g., analytic). Viewing constraint sets from this angle and following the pioneering work [47], we establish here various genericity results for regular perturbations.

Smooth constraint functions. The first genericity result we present here concerns a *linear independence constraint qualification*, a strong and quite stringent qualification condition which is often considered in the literature when dealing with "generic" instances of optimization problems (see, e.g., [36, 41]). This qualification condition requires that the gradients of both the equality constraints and the active inequality constraints are linearly independent. Note that it implies in particular MFCQ.

Let us first recall the classical Sard theorem. For a differentiable map $f : \mathbb{R}^p \to \mathbb{R}^q$, a point $x \in \mathbb{R}^p$ is *critical* if the differential mapping of f at x is not surjective. A *critical value* of f is the image of a critical point. Otherwise, v is said to be *regular*.

SARD'S THEOREM 2.6 (see [38]). *Let $f : \mathbb{R}^p \to \mathbb{R}^q$ be a map of class $\mathcal{C}^k$ with $k > \max(0, p-q)$. Then the Lebesgue measure of the set of critical values of f is zero.*

As a consequence of Sard's theorem, we deduce that a perturbation of the constraint set of problem $(\mathcal{P}_{\text{NLP}})$ is almost surely regular when the constraint functions are smooth enough.

THEOREM 2.7 (compare with [47, Thm. 1]). *Let $g_1, \dots, g_m, h_1, \dots, h_r$ be $\mathcal{C}^k$ constraint functions from $\mathbb{R}^n$ to $\mathbb{R}$ with $k > \max(0, n-r)$. Then the set of admissible perturbations $(\mu, \nu) \in \mathbb{R}^m \times \mathbb{R}^r$ for which the linear independence constraint qualification is not satisfied at every point of the set $C_\mu \cap M_\nu$ has Lebesgue measure zero. In particular, the set $\mathcal{A}_{\text{sing}}$ of singular perturbations has Lebesgue measure zero.*

Definable constraint functions. The above result can be considerably relaxed by replacing smoothness assumptions by mere definability. The results on definability and tame geometry that we use hereafter are recalled in Appendix A.

Ioffe showed a nonsmooth version of Sard's theorem for definable set-valued mappings. In this framework, a vector $\bar{y} \in \mathbb{R}^q$ is a *critical value* of any set-valued mapping $F : \mathbb{R}^p \rightrightarrows \mathbb{R}^q$ if there exists a point $\bar{x} \in \mathbb{R}^p$ such that $\bar{y} \in F(\bar{x})$ and F is not metrically regular at $(\bar{x}, \bar{y})$.

NONSMOOTH SARD'S THEOREM 2.8 (see [32, Thm. 1]). *Let $F : \mathbb{R}^p \rightrightarrows \mathbb{R}^q$ be a definable set-valued mapping with locally closed graph. Then the set of critical values of F is a definable set in $\mathbb{R}^q$ whose dimension is less than $q-1$.*

Combining this result with Theorem 2.5, we readily get a geometric description of regular perturbations for problem $(\mathcal{P}_{\text{NLP}})$ when the constraint functions are definable in the same o-minimal structure. Let us mention that we also use the fact that any definable set $A \subset \mathbb{R}^p$ can be "stratified," that is, written as a finite disjoint union of smooth submanifolds of $\mathbb{R}^p$ that fit together in a "regular" manner. This implies in particular that the dimension of A, i.e., the largest dimension of such submanifolds, is strictly lower than p if and only if the complement of A is dense.

THEOREM 2.9 (genericity of regular perturbations). *Let $g_1, \dots, g_m, h_1, \dots, h_r : \mathbb{R}^n \to \mathbb{R}$ be constraint functions that are definable in the same o-minimal structure.*

Then the set $\mathcal{A}_{\text{reg}}$ (resp., $\mathcal{A}_{\text{sing}}$) of regular (resp., singular) perturbations is definable in $\mathbb{R}^{m+r}$ and $\mathcal{A}_{\text{sing}}$ is a finite union of smooth submanifolds of $\mathbb{R}^{m+r}$ of dimension strictly lower than $m+r$.

Remark 2.10. Note that, in general, the set $\mathcal{A}_{\text{sing}}$ of singular perturbations is not closed. Consider for instance the semialgebraic functions defined on $\mathbb{R}^2$ by

$$g_1(x_1,x_2) = \min\left\{2x_1 - 1, \frac{1}{|x_1|}\right\} - x_2,$$
$$h_1(x_1,x_2) = \frac{x_1}{1+(x_1)^2} - x_2, \quad h_2(x_1,x_2) = \frac{x_1}{1+(x_1)^2} + x_2.$$

For every $\nu \in (0,\frac{1}{2})$, the set $[h_1 = \nu, h_2 = \nu]$ contains two distinct points, namely $(\frac{1\pm\sqrt{1-4\nu^2}}{2\nu}, 0)$, whereas $[h_1 = 0, h_2 = 0] = \{(0,0)\}$. Let $\mu_\nu = \frac{1+\sqrt{1-4\nu^2}}{2\nu}$. One easily checks that, for the constraint set $[g_1 \leqslant \mu_\nu^{-1}, h_1 = \nu, h_2 = \nu]$ with $\nu \in (0,\frac{1}{2})$, MFCQ fails at point $(\mu_\nu, 0)$ but it is satisfied at point $(\frac{1-\sqrt{1-4\nu^2}}{2\nu}, 0)$ where the inequality constraint is not active. Hence (μ_ν^{-1}, ν, ν) is a singular perturbation for all $\nu \in (0,\frac{1}{2})$. However, when $\nu = 0$ the singularity disappears and only the point $(0,0)$ remains, at which MFCQ holds. In other words, $(0,0,0)$ is regular.

2.4. Continuity properties of perturbations. We investigate below the continuity properties of the perturbed constraint sets and of the value function of problem $(\mathcal{P}_{\text{NLP}})$. Recall beforehand that given any set-valued mapping $F : \mathbb{R}^p \rightrightarrows \mathbb{R}^q$, the *outer limit*, $\limsup_{x\to\bar{x}} F(x) \subset \mathbb{R}^q$, and the *inner limit*, $\liminf_{x\to\bar{x}} F(x) \subset \mathbb{R}^q$, of F at any point $\bar{x} \in \mathbb{R}^p$ are respectively defined by the following:

$$y \in \limsup_{x\to\bar{x}} F(x) \iff \exists\, x_n \to \bar{x}, \quad \exists\, y_n \to y \quad \forall\, n \in \mathbb{N}, \quad y_n \in F(x_n),$$
$$y \in \liminf_{x\to\bar{x}} F(x) \iff \forall\, x_n \to \bar{x}, \quad \exists\, y_n \to y, \quad \exists\, n_0 \in \mathbb{N} \quad \forall\, n \geqslant n_0, \quad y_n \in F(x_n).$$

Then we can define the notion of (semi)continuity for set-valued mappings.

DEFINITION 2.11 (semicontinuity of set-valued mappings). *A set-valued mapping $F : \mathbb{R}^p \rightrightarrows \mathbb{R}^q$ is* outer semicontinuous *(resp.,* inner semicontinuous*) at $\bar{x} \in \mathbb{R}^p$ if*

$$\limsup_{x\to\bar{x}} F(x) \subset F(\bar{x}) \quad \Big(\textit{resp.,}\ \liminf_{x\to\bar{x}} F(x) \supset F(\bar{x})\Big).$$

It is continuous *at $\bar{x}$ if it is both outer and inner semicontinuous.*

A straightforward application of these definitions leads to the following elementary lemma.

LEMMA 2.12 (continuity of perturbed sets). *Let $g_1,\ldots,g_m : \mathbb{R}^n \to \mathbb{R}$ be continuous functions. Assume that the constraint set $C_0 = [g_1 \leqslant 0,\ldots,g_m \leqslant 0]$ is nonempty. Then the set-valued mapping $\mathbb{R}^m_+ \rightrightarrows \mathbb{R}^n$, $\mu \mapsto C_\mu$ is continuous at* 0. □

Remark 2.13. It is in general necessary to consider nonnegative perturbations in order to have continuity at 0. Indeed, for general perturbations, although the inequality constraint set mapping $\mathbb{R}^m \rightrightarrows \mathbb{R}^n$, $\mu \mapsto C_\mu$ is outer semicontinuous at 0 (this readily follows from the continuity of the constraint functions), it is not inner semicontinuous. Consider for instance the following constraint set, defined for any $\mu \in \mathbb{R}^2$ by

$$C_\mu = \{x \in \mathbb{R} \mid 1 - x^2 \leqslant \mu_1,\ (x+1)^2 - 4 \leqslant \mu_2\}$$

and check that $C_0 = [-3,-1] \cup \{1\}$. However, for all $\mu_1 < 0$ and $\mu_2 < 0$ small enough, we have $C_\mu = [-1-\sqrt{4+\mu_2}, -\sqrt{1-\mu_1}] \subset [-3,-1]$. Hence $\{1\}$ cannot be in the inner limit of C_μ as $\mu \to 0$. Precisely, we have $\liminf_{\mu\to 0} C_\mu = [-3,-1]$.

As for the equality constraint set mapping $\mathbb{R}^r \rightrightarrows \mathbb{R}^n$, $\nu \mapsto M_\nu$, it is also clearly outer semicontinuous at 0 but not inner semicontinuous in general, even when restricted to $\mathbb{R}^r_+$. For instance, consider for $\nu \in \mathbb{R}$ the constraint set

$$M_\nu = \{x \in \mathbb{R} \mid 9x(x^2-1) - 2\sqrt{3} = \nu\}$$

and check that $M_0 = \{-1/\sqrt{3},\ 2/\sqrt{3}\}$. However, for every $\nu > 0$, M_ν contains a unique point, which converges to $2/\sqrt{3}$ as ν tends to 0. As a consequence, the inner limit of M_ν at 0 can only contain this point. Studying the situation when $\nu < 0$, one readily see that, actually, $\liminf_{\nu\to 0} M_\nu = \{2/\sqrt{3}\}$.

We now turn our attention to the behavior of the value function of perturbed problems and we study the continuity at 0 of $(\mu,\nu) \mapsto \min\{f(x) \mid x \in C_\mu \cap M_\nu\}$. As a consequence of previous observations, continuity cannot occur in general when equality constraints are present, and it is "necessary" to consider nonnegative perturbations for the inequalities. The next result is a classical one; see, e.g., [14, Prop. 4.4]. In the following, we denote by $\mathbb{R}^m_{++}$ the set of vectors in $\mathbb{R}^m$ with positive entries.

LEMMA 2.14 (continuity of the value function). *Let $f, g_1, \ldots, g_m : \mathbb{R}^n \to \mathbb{R}$ be continuous functions. Assume that the constraint set $C_\mu = [g_1 \leqslant \mu_1, \ldots, g_m \leqslant \mu_m]$ is nonempty for $\mu = 0$ and bounded for some positive perturbation $\mu' \in \mathbb{R}^m_{++}$. Then the value function* $\mathrm{val} : \mathbb{R}^m_+ \to \mathbb{R}$ *defined by* $\mathrm{val}(\mu) = \min_{x\in C_\mu} f(x)$ *is continuous at* 0*:*

$$\min_{x\in C_\mu} f(x) \xrightarrow[\mu\in\mathbb{R}^m_+]{\mu\to 0,} \min_{x\in C_0} f(x).$$

Proof. First, since $C_0 \subset C_\mu$ for all $\mu \in \mathbb{R}^m_+$, we have $\mathrm{val}(0) \geqslant \limsup_{\mu\to 0} \mathrm{val}(\mu)$.

Let $(\mu_n)_{n\in\mathbb{N}}$ be any sequence in $\mathbb{R}^m_+$ converging to 0 and such that $\mathrm{val}(\mu_n)$ converges to some $v \in \mathbb{R} \cup \{-\infty\}$. Since $\mu' \in \mathbb{R}^m_{++}$, we may assume without loss of generality that we have, for all integers n, $\mu' - \mu_n \in \mathbb{R}^m_+$, so that $C_{\mu_n} \subset C_{\mu'}$. Let $x^*_n \in \arg\min\{f(x) \mid x \in C_{\mu_n}\}$, that is, $x^*_n \in C_{\mu_n}$ and $f(x^*_n) = \mathrm{val}(\mu_n)$. Since the sequence $(x^*_n)_{n\in\mathbb{N}}$ lies in the bounded set $C_{\mu'}$, it converges, up to an extraction, to some point x^*. Therefore, we have $v = f(x^*)$ with $x^* \in C_0$ by continuity of f and $\mu \mapsto C_\mu$ (Lemma 2.12). We deduce that $\mathrm{val}(0) \leqslant \liminf_{\mu\to 0} \mathrm{val}(\mu)$, which concludes the proof. ∎

Note that, without the compactness assumption, the conclusion of Lemma 2.14 does not hold. Consider for instance the semialgebraic programming problem

$$\text{minimize } \frac{1+x^2}{1+x^4} \quad \text{subject to } x \in \mathbb{R},\ \frac{x^2}{1+x^4} \leqslant \alpha.$$

For all scalars $\alpha > 0$, the value of the problem is $\mathrm{val}(\alpha) = 0$, whereas for $\alpha = 0$ it is $\mathrm{val}(0) = 1$.

3. Finiteness of singular diagonal perturbations.

3.1. Geometric aspects of regular perturbations. Although Theorem 2.9 is a satisfying theoretical result, it does not give any structural information beyond dimension and definability. In particular it is not clear *how the perturbations* should be chosen when dealing with concrete optimization problems. The following result shows that, under reasonable assumptions, *small positive* and *small negative* perturbations μ of the inequality constraints are always regular, that is, MFCQ is satisfied at every point in C_μ. For the sake of simplicity, we have chosen to state this result as well as all the subsequent ones for constraint sets defined only by inequalities. Nevertheless, they all extend easily to the setting of inequality and equality constraints (with perturbations applying only to the inequalities); see Remark 3.2(b) and Remarks 3.4, 3.6, 3.8 and 3.11.

THEOREM 3.1 (small regular perturbations). *Let $g_1, \dots, g_m : \mathbb{R}^n \to \mathbb{R}$ be differentiable constraint functions that are definable in the same o-minimal structure.*

(i) Outer regular perturbations. *If $C_0 = [g_1 \leqslant 0, \dots, g_m \leqslant 0]$ is nonempty, then there exists $\varepsilon_0 > 0$ such that $(0, \varepsilon_0)^m \subset \mathcal{A}_{\mathrm{reg}}$. In other words, for all positive perturbations $\mu \in (0, \varepsilon_0)^m$, the Mangasarian–Fromovitz constraint qualification holds throughout $C_\mu = [g_1 \leqslant \mu_1, \dots, g_m \leqslant \mu_m]$.*

(ii) Inner regular perturbations. *If $[g_1 < 0, \dots, g_m < 0]$ is nonempty, then there exists $\varepsilon_1 > 0$ such that $(-\varepsilon_1, 0)^m \subset \mathcal{A}_{\mathrm{reg}}$. In other words, for all negative perturbations $\mu \in (-\varepsilon_1, 0)^m$, the Mangasarian–Fromovitz constraint qualification holds throughout $C_\mu = [g_1 \leqslant \mu_1, \dots, g_m \leqslant \mu_m]$.*

Proof. We only show (i). Item (ii) follows from very similar arguments. Let us first notice that the constraint set mapping $\mathbb{R}^m_+ \rightrightarrows \mathbb{R}^n$, $\mu \mapsto C_\mu$ is a definable mapping. For each μ in $\mathbb{R}^m_{++}$ we consider the subset $S(\mu)$ of C_μ consisting of the points at which MFCQ is not satisfied. Following Remark 2.2, we have

$$S(\mu) = \Big\{ x \in C_\mu \mid 0 \in \mathrm{co}\,\{\nabla g_i(x) \mid i \in I(x)\} \Big\}. \tag{3.1}$$

This extends to a definable set-valued mapping $S : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$, $\mu \mapsto S(\mu)$ by setting $S(\mu) = \emptyset$ if $\mu \notin \mathbb{R}^m_{++}$.

Working towards a contradiction, we assume that 0 belongs to the closure of $\operatorname{dom} S$. Using Curve Selection Lemma A.6, we obtain a definable $\mathcal{C}^1$ curve $[0,1) \to \mathbb{R}^m$, $t \mapsto \mu(t)$ such that $\mu(t) \in \operatorname{dom} S$ for all $t > 0$ and $\mu(0) = 0$. Monotonicity Lemma A.4 combined with the fact that $\mu(t) \in \operatorname{dom} S \subset \mathbb{R}^m_{++}$ for $t \in (0,1)$ and $\mu(0) = 0$ ensures the existence of $\varepsilon > 0$ such that

$$\dot{\mu}_i(t) > 0, \quad i = 1, \dots, m, \quad \forall t \in (0, \varepsilon). \tag{3.2}$$

The set-valued mapping $(0, \varepsilon) \rightrightarrows \mathbb{R}^n$, $t \mapsto S(\mu(t))$ is definable and has nonempty values, hence Definable Choice Lemma A.5 yields the existence of a definable curve $x : (0, \varepsilon) \to \mathbb{R}^n$ such that $x(t) \in S(\mu(t))$ for all t. Shrinking ε if necessary (using Monotonicity Lemma A.4) we can assume that $x(\cdot)$ is $\mathcal{C}^1$.

Given two definable functions $a, b : (0, \varepsilon) \to \mathbb{R}$, we can apply once more Monotonicity Lemma A.4 to see that either $a(t) = b(t)$ or $a(t) > b(t)$ or $b(t) > a(t)$ for t sufficiently small. This implies in particular that there exists a positive real $\varepsilon' \leqslant \varepsilon$ and a nonempty subset $I \subset \{1, \dots, m\}$ such that $I(x(t)) = I$ for all $t \in (0, \varepsilon')$. Indeed, recall that $I(x(t)) = \{1 \leqslant i \leqslant m \mid g_i(x(t)) = \mu_i(t)\}$, where each pair of functions $(g_i(x(\cdot)), \mu_i(\cdot))$, $i = 1, \dots, m$, is definable. Hence $I(x(t))$ stabilizes for $t > 0$ sufficiently small. Furthermore, for all $t \in (0, \varepsilon)$, $I(x(t))$ is nonempty because otherwise MFCQ would be satisfied at $x(t)$, which would contradict the fact that $x(t) \in S(\mu(t))$.

By definition of S, for all $t \in (0, \varepsilon')$ there exist coefficients $\lambda_i(t)$ with $i \in I$ such that

$$\lambda_i(t) \geqslant 0, \ \forall i \in I, \quad \text{and} \quad \sum_{i \in I} \lambda_i(t) = 1,$$

and such that

$$\sum_{i\in I} \lambda_i(t)\,\nabla g_i(x(t)) = 0. \tag{3.3}$$

Multiplying each member of the above equality by $\dot{x}(t)$, one obtains

$$\sum_{i\in I} \lambda_i(t)\,\langle\, \dot{x}(t), \nabla g_i(x(t))\,\rangle = 0,$$

which can also be written

$$\sum_{i\in I} \lambda_i(t)\,\frac{d(g_i \circ x)}{dt}(t) = 0.$$

Since each inequality constraint g_i, $i \in I$, is active for all $t \in (0, \varepsilon')$, one gets

$$\sum_{i\in I} \lambda_i(t)\,\dot{\mu}_i(t) = 0, \quad t \in (0, \varepsilon'),$$

a contradiction: indeed (3.2) and the fact that $I \neq \emptyset$ indicate that the left-hand side of the latter equality is positive for all $t \in (0, \varepsilon')$.

For (ii), it suffices to notice that Slater's condition guarantees that the set C_μ with $\mu \in (-\infty, 0)^m$ is nonempty for μ in a neighborhood of 0. The proof then follows arguments similar to those developed for (i). ∎

We next discuss some aspects of the hypotheses of Theorem 3.1.

Remark 3.2. (a) A similar result cannot be derived for joint perturbations (μ, ν) of inequality and equality constraint sets. Consider for instance the functions defined on $\mathbb{R}^2$ by $g(x_1, x_2) = x_2$ and $h(x_1, x_2) = x_2 - (x_1)^2$. For all perturbations $\mu \in \mathbb{R}$, the constraint set $C_\mu \cap M_\mu = [g \leqslant \mu, h = \mu]$ contains only the point $(0, \mu)$, at which MFCQ does not hold since $\nabla g(0, \mu) = \nabla h(0, \mu) = (0, 1)$. (See also Remark 2.10.)

(b) The conclusion of Theorem 3.1 still holds for perturbed constraint sets of the form $C_\mu \cap M_0$. For this, one needs to assume that the equality constraint set $M_0 = [h_1 = 0, \ldots, h_r = 0]$ is defined by definable differentiable functions $h_1, \ldots, h_r : \mathbb{R}^n \to \mathbb{R}$ whose gradient vectors $\nabla h_j(x)$, $j = 1, \ldots, r$, are linearly independent for all $x \in M_0$.

Sketch of proof. Only a few changes in the proof of the previous theorem are necessary. The first one is the definition of the set-valued map S (see (3.1)), which sends perturbation vectors $\mu \in \mathbb{R}^m$ to the set of feasible points at which constraint qualification conditions are not satisfied. In this new setting, it becomes

$$S(\mu) = \Big\{ x \in C_\mu \cap M_0 \mid \mathrm{co}\,\{\nabla g_i(x) \mid i \in I(x)\} \cap \mathrm{span}\,\{\nabla h_j(x) \mid 1 \leqslant j \leqslant r\} \neq \emptyset \Big\}.$$

The second change is (3.3), which characterizes the failure of the constraint qualification at point $x(t) \in S(\mu(t))$. Since the gradient vectors of the equality constraints are linearly independent for all the feasible points, this failure of constraint qualification must come from the absence of a vector y satisfying (2.1). Hence, following Remark 2.2, the right-hand side of the equality must be replaced by a linear combination of the gradients ∇h_j at point $x(t)$, that is, (3.3) now reads

$$\sum_{i\in I} \lambda_i(t)\,\nabla g_i(x(t)) = \sum_{j=1}^{r} \kappa_j(t)\,\nabla h_j(x(t))$$

for some coefficients $\kappa_j(t) \in \mathbb{R}$, $j = 1, \ldots, r$. Then the proof proceeds along the same lines. In particular, multiplying the right-hand side of the previous equality by $\dot{x}(t)$, one obtains

$$\sum_{j=1}^{r} \kappa_j(t) \langle \dot{x}(t), \nabla h_j(x(t)) \rangle = \sum_{j=1}^{r} \kappa_j(t) \frac{d(h_j \circ x)}{dt}(t) = 0$$

since each equality constraint h_j, $j = 1, \ldots, r$, is constant on the curve x.

(c) The use of small perturbation vectors which are not positive (or negative) may not remove the absence of MFCQ. A simple example is given on $\mathbb{R}^2$ by

$$g_1(x_1, x_2) = (x_1)^2 + (x_2)^2 - 1 \quad \text{and} \quad g_2(x_1, x_2) = (x_1 - 2)^2 + (x_2)^2 - 1.$$

The sets $[g_1 \leqslant 0]$ and $[g_2 \leqslant 0]$ delineate two tangent discs. Therefore, MFCQ fails at their contact point $(1, 0)$. Consider the perturbation path $t \mapsto (\mu_1(t), \mu_2(t)) = (t^2 - 2t, t^2 + 2t)$ which passes through $(0, 0)$ with velocity $(2, -2)$. It can be checked that the constraint set $[g_1 \leqslant \mu_1(t), g_2 \leqslant \mu_2(t)]$ is not regular at $(1 - t, 0)$ for all $t \in (-1, 1)$.

(d) Even though definable functions are not the unique class of functions for which a theorem similar to Theorem 3.1 can be derived,[3] the definability assumption in Theorem 3.1 cannot be replaced by mere smoothness. Many counterexamples can be given, even when $n = 1$. Consider for instance the strictly increasing $\mathcal{C}^\infty$ function $g(x) = \int_0^x \exp(-t^{-2}) \sin^2(t^{-1}) dt$ with $x \in \mathbb{R}$, and the set $[g \leqslant 0] = (-\infty, 0]$. Obviously the set of regular perturbations $\mathcal{A}_{\text{reg}}$ does not contain any segment of the form $(0, \varepsilon)$ with $\varepsilon > 0$.

We now provide a "partial perturbation version" of our main result, which can be proved following the lines of Theorem 3.1. It relies on the assumption that the set defined by the first p inequalities is regular.

THEOREM 3.3 (partial constraint qualification). *Let $g_1, \ldots, g_m : \mathbb{R}^n \to \mathbb{R}$ be differentiable functions that are definable in the same o-minimal structure. Assume that $C_0 = [g_1 \leqslant 0, \ldots, g_m \leqslant 0]$ is nonempty and that the Mangasarian–Fromovitz constraint qualification holds throughout $[g_1 \leqslant 0, \ldots, g_p \leqslant 0]$ for some positive integer $p < m$. Then there exists $\varepsilon > 0$ such that, for all perturbations $\mu_{p+1}, \ldots, \mu_m \in (0, \varepsilon)$, MFCQ holds throughout $[g_1 \leqslant 0, \ldots, g_p \leqslant 0, g_{p+1} \leqslant \mu_{p+1}, \ldots, g_m \leqslant \mu_m]$.* ∎

Remark 3.4. Similarly to Remark 3.2(b), Theorem 3.3 also holds in the setting of fixed equality constraints, in addition to partially perturbed inequality constraints.

A simple but important corollary is a kind of partial Slater qualification condition.

COROLLARY 3.5 (partial Slater condition). *Let $g_1, \ldots, g_m : \mathbb{R}^n \to \mathbb{R}$ be differentiable convex functions that are definable in the same o-minimal structure. Assume that $C_0 \neq \emptyset$ and that $g_i(x_0) < 0$ for some $x_0 \in \mathbb{R}^n$ and all $i \in \{1, \ldots, p\}$, where $p < m$. Then there exists $\varepsilon > 0$ such that, for all perturbations $\mu_{p+1}, \ldots, \mu_m \in (0, \varepsilon)$, MFCQ holds throughout $[g_1 \leqslant 0, \ldots, g_p \leqslant 0, g_{p+1} \leqslant \mu_{p+1}, \ldots, g_m \leqslant \mu_m]$.* ∎

Remark 3.6. The above result is provided without equality constraints for the sake of simplicity: the addition of a finite system of affine constraints[4] is an easy task.

[3] One can think for instance of continuous convex functions.

[4] The linear independence assumption of Remark 3.2(b) is not necessary in this case.

To conclude this subsection, let us emphasize that Theorem 3.1 applies to several frameworks that are widely spread in practice, including polynomial, semialgebraic, real analytic, and many other kinds of definable constraints.

3.2. Diagonal perturbations. When the constraint functions are definable in the same o-minimal structure, then so are the sets of regular and singular perturbations, since they can be described by a first-order formula (see also Theorem 2.9). Thus, a direct application of Theorem 3.1 leads to the finiteness of singular perturbations along any direction, and in particular along the *diagonal*, i.e., for perturbations of the form $(\alpha, \ldots, \alpha)$ with $\alpha \in \mathbb{R}$. With a minor abuse of notation, for $\alpha \in \mathbb{R}$, we use $C_\alpha := [g_1 \leqslant \alpha, \ldots, g_m \leqslant \alpha]$.

COROLLARY 3.7. *Let $g_1, \ldots, g_m : \mathbb{R}^n \to \mathbb{R}$ be differentiable functions that are definable in the same o-minimal structure. Then, for all except finitely many perturbations $\alpha \in \mathbb{R}$, the Mangasarian–Fromovitz constraint qualification holds throughout $C_\alpha = [g_1 \leqslant \alpha, \ldots, g_m \leqslant \alpha]$. The same conclusion holds for*

$$[g_1 \leqslant 0, \ldots, g_p \leqslant 0, g_{p+1} \leqslant \alpha, \ldots, g_m \leqslant \alpha]$$

if MFCQ holds throughout $[g_1 \leqslant 0, \ldots, g_p \leqslant 0]$ for some $p < m$.

Remark 3.8 (equality and inequality constraint). Similarly to Remark 3.2(b), the above result also holds for perturbed constrained sets of the form $C_\alpha \cap M_0$, where M_0 is defined by definable functions that are differentiable and whose gradients are independent at any point $x \in M_0$.

Remark 3.9 (diagonal perturbations through the nonsmooth Sard theorem). With a slightly stronger regularity assumption, Corollary 3.7 can be seen as the nonsmooth definable Sard-type theorem [10, Cor. 9]. We next explain this observation.

When dealing with diagonal perturbations $\alpha \in \mathbb{R}$, the constraint set C_α can be represented as the lower level set of a single real-extended-valued function. Namely, a point x is in C_α if and only if

$$\max_{1 \leqslant i \leqslant m} g_i(x) \leqslant \alpha.$$

Let us define $g = \max_{1 \leqslant i \leqslant m} g_i$ and let us assume that the constraint functions $g_1, \ldots, g_m$ are $\mathcal{C}^1$. This implies that the basic chain rule of subdifferential calculus applies to the function g (see [43, Thm. 10.6]). Thus we have, for all $x \in \mathbb{R}^n$,

$$\widehat{\partial} g(x) = \partial g(x) = \operatorname{co} \{\nabla g_i(x) \mid 1 \leqslant i \leqslant m,\ g_i(x) = g(x)\},$$

where, for any function $f : \mathbb{R}^n \to \mathbb{R}$ and any $x \in \mathbb{R}^n$, $\widehat{\partial} f(x)$ and $\partial f(x)$ respectively denote the *Fréchet subdifferential* and the *limiting/Mordukhovich subdifferential* of f at x; see [43] for their constructions.

Now observe that $\alpha \in \mathbb{R}$ is a singular perturbation, that is, MFCQ is not satisfied at some point $x \in C_\alpha$, if and only if $g(x) = \alpha$ and $0 \in \partial g(x)$. In other words, the singular diagonal perturbations of the inequality constraint set of problem $(\mathcal{P}_{\text{NLP}})$ correspond to the *critical values* of g.

Thus, when the functions $g_1, \ldots, g_m$ are definable in the same o-minimal structure, so is g, and the finiteness of the singular diagonal perturbations stated in Corollary 3.7 is equivalent to the finiteness of the critical values of the definable function g. Hence, with continuously differentiable functions, Corollary 3.7 can be seen as a Sard-type theorem for definable functions, which was proved in full generality in [10, Cor. 9]. These arguments can also be extended when equality constraints with linearly independent gradients are added to the constraint set (see Remark 3.2(b)).

3.3. A bound on the number of singular perturbations for polynomial optimization. We consider here constraint sets defined by real polynomial functions and we bound the number of singular values for the corresponding perturbed sets.

To tackle this problem, we evaluate the number of connected components of some adequate real algebraic sets. A key result regarding this evaluation is provided by the Milnor–Thom bound: given any polynomial map $f : \mathbb{R}^p \to \mathbb{R}^q$, the number of connected components of the set of zeros of f, $\{x \in \mathbb{R}^p \mid f(x) = 0\}$, is bounded by

$$d\,(2d-1)^{p-1}, \tag{3.4}$$

where d is the maximal degree of the polynomial functions f_j for $j = 1, \ldots, q$ (see, e.g., [4]).

THEOREM 3.10. *Let $g_1, \ldots, g_m : \mathbb{R}^n \to \mathbb{R}$ be polynomial functions whose degrees are bounded by d. Let $\{I, J\}$ be a partition of the set of indices $\{1, \ldots, m\}$, possibly trivial, such that the Mangasarian–Fromovitz constraint qualification holds throughout $[g_j \leqslant 0,\, j \in J]$. Then, for the perturbed sets $[g_i \leqslant \alpha, g_j \leqslant 0, i \in I, j \in J]$ with α ranging in $\mathbb{R}$, the number of singular perturbations is bounded by*

$$d\,(2d-1)^n\,(2d+1)^m.$$

Proof. Denoting by $|J|$ the cardinality of J, we assume that $|J| \in \{0, \ldots, m-1\}$ (otherwise I is empty and there is nothing to prove). For $\alpha \in \mathbb{R}$, let

$$C_{I,\alpha} := [g_i \leqslant \alpha, g_j \leqslant 0,\, i \in I,\, j \in J].$$

If α is a singular perturbation, then there exists a point $x \in C_{I,\alpha}$ for which $0 \in \operatorname{co}\{\nabla g_i(x) \mid i \in I(x)\}$. So, there exists a subset of indices $K \subset I(x)$, which we fix, and positive scalars $\lambda_i > 0$, $i \in K$, such that $\sum_{i\in K} \lambda_i \nabla g_i(x) = 0$ and $\sum_{i\in K} \lambda_i = 1$. Furthermore, $K \not\subset J$ since MFCQ holds throughout $[g_j \leqslant 0, j \in J]$. Let L be the set of indices equal to $I(x)$. Thus, the sets K and L being fixed, the tuple $(x, \lambda, \alpha) \in \mathbb{R}^n \times \mathbb{R}^K \times \mathbb{R}$ is a solution of the polynomial system

$$\begin{cases} \displaystyle\sum_{i\in K} \lambda_i \nabla g_i(x) = 0, \\ \displaystyle\sum_{i\in K} \lambda_i = 1, \\ g_j(x) = \alpha, \quad j \in L \cap I, \\ g_j(x) = 0, \quad j \in L \cap J, \end{cases} \tag{3.5}$$

and satisfies the following additional constraints: $\lambda \in \mathbb{R}^K_{++}$, $g_\ell(x) < \alpha$ for all $\ell \in I \setminus L$, and $g_\ell(x) < 0$ for all $\ell \in J \setminus L$.

The first step of the proof is to show that the number of singular perturbations is bounded above by the number of connected components of the set of solutions of (3.5) for all possible choices of K and L. This is done by constructing an injection from the set of singular perturbations to these connected components.

Fix a singular value α and choose a subset $L \subset \{1, \ldots, m\}$ with maximal cardinality among all the sets of active constraints $I(x)$ such that MFCQ is not satisfied at $x \in C_{I,\alpha}$. Then pick a subset $K \subset L$ with minimal cardinality among all the subsets $K' \subset L$ such that the system (3.5) with K replaced by K' has a solution $(x, \lambda, \alpha) \in \mathbb{R}^n \times \mathbb{R}^{K'} \times \mathbb{R}$ with $x \in C_{I,\alpha}$ and $\lambda \in \mathbb{R}^{K'}_+$. Let $(\bar{x}, \bar{\lambda}, \alpha)$ be such a solution for the particular choice of K and L. Note that $K \not\subset J$ since $[g_j \leqslant 0, j \in J]$ is regular. Also note that $\bar{\lambda} \in \mathbb{R}^K_{++}$ by minimality of $|K|$, and that $g_\ell(\bar{x}) < \alpha$ for all $\ell \in I \setminus L$, and $g_\ell(\bar{x}) < 0$ for all $\ell \in J \setminus L$ by maximality of $|L|$.

Let $Q \subset \mathbb{R}^n \times \mathbb{R}^K \times \mathbb{R}$ be the connected component of the set of solutions of (3.5) corresponding to K and L, containing the tuple $(\bar{x}, \bar{\lambda}, \alpha)$. We next prove that

$$Q \subset S(\alpha, K, L) := \{x \in C_{I,\alpha} \mid I(x) = L\} \times \mathbb{R}^K_{++} \times \{\alpha\}.$$

Working towards a contradiction, assume that the above inclusion does not hold. There exists therefore a continuous path $(x(\cdot), \lambda(\cdot), \alpha(\cdot))$ from $[0,1]$ to Q such that

$$\begin{cases} (x(0), \lambda(0), \alpha(0)) = (\bar{x}, \bar{\lambda}, \alpha), \\ (x(1), \lambda(1), \alpha(1)) \notin S(\alpha, K, L). \end{cases}$$

Let $t = \sup\{s \in [0,1] \mid (x(s), \lambda(s), \alpha(s)) \in S(\alpha, K, L)\}$. By continuity we have $\alpha(t) = \alpha$, $x(t) \in C_{I,\alpha}$, and $\lambda(t) \in \mathbb{R}^K_+$. If either $I(x(t)) \neq L$ or $\lambda(t) \notin \mathbb{R}^K_{++}$, then there is a contradiction with the maximality of $|L|$ (since we already have $L \subset I(x(t))$) or with the minimality of $|K|$. Hence, we have $I(x(t)) = L$ and $\lambda(t) \in \mathbb{R}^K_{++}$. Since in addition $x(t) \in C_{I,\alpha}$ and $\alpha(t) = \alpha$, we have $(x(t), \lambda(t), \alpha(t)) \in S(\alpha, K, L)$. Finally, we have $t < 1$ since $(x(1), \lambda(1), \alpha(1)) \notin S(\alpha, K, L)$. Using the continuity of the path and (3.5), there exists $\varepsilon > 0$ such that for all $s \in [t, t+\varepsilon)$, we have $x(s) \in C_{I,\alpha(s)}$ with $I(x(s)) = L$ and $\lambda(s) \in \mathbb{R}^K_{++}$. This implies that $\alpha(s)$ is a singular perturbation for all $s \in [t, t+\varepsilon)$. Combining the continuity of $\alpha(\cdot)$ and Corollary 3.7, $\alpha(\cdot)$ is constant on $[t, t+\varepsilon)$. Hence, we have $(x(s), \lambda(s), \alpha(s)) \in S(\alpha, K, L)$ for all $s \in [t, t+\varepsilon)$. From the definition of t, we obtain $t \geqslant t + \varepsilon$, which is contradictory since $\varepsilon > 0$.

Thus, for every singular perturbation α, there exist subsets $K \subset L \subset \{1, \ldots, m\}$ with $K \cap I \neq \emptyset$ such that the set of solutions of the polynomial system (3.5) with this choice of K and L has at least one connected component included in $\mathbb{R}^n \times \mathbb{R}^K \times \{\alpha\}$. Hence the mapping sending every singular perturbation to this connected component is injective. So we have just proved that the number of singular perturbations is upper bounded by the number of connected components of the set of solutions of (3.5) for all possible choices of K and L. We can then deduce from the Milnor–Thom bound (3.4) an upper bound for the number of singular perturbations α by summation over all possible choices of K and L.

Define $p = |I| \in \{1, \ldots, m\}$. In the computation below we denote by ℓ_1, ℓ_2 the cardinality of $L \cap I$ and $L \cap J$, respectively, and by k_1, k_2 the cardinality of $K \cap I$ and $K \cap J$, respectively. Since the system (3.5) has degree d and $n + k_1 + k_2 + 1$ variables, the number of singular perturbation is bounded by

$$\begin{aligned} &\sum_{\substack{1 \leqslant \ell_1 \leqslant p \\ 0 \leqslant \ell_2 \leqslant m-p}} \binom{p}{\ell_1} \binom{m-p}{\ell_2} \sum_{\substack{1 \leqslant k_1 \leqslant \ell_1 \\ 0 \leqslant k_2 \leqslant \ell_2}} \binom{\ell_1}{k_1} \binom{\ell_2}{k_2} d\,(2d-1)^{n+k_1+k_2} \\ &\qquad = d\,(2d-1)^n\,(2d+1)^{m-p} \left((2d+1)^p - 2^p\right) \\ &\qquad = d\,(2d-1)^n\,(2d+1)^m \left(1 - \left(\frac{2}{2d+1}\right)^p\right). \end{aligned}$$

To conclude, observe that

$$\frac{1}{3} \leqslant 1 - \left(\frac{2}{2d+1}\right)^p \leqslant 1 \tag{3.6}$$

for all $d \geqslant 1$ and $1 \leqslant p \leqslant m$. ∎

Remark 3.11. (a) As demonstrated in a forthcoming example, the choice of a partition (I, J) has a very marginal impact on the global bound, which we have neglected in our main estimate (3.4).[5]

(b) Let $h_1, \dots, h_r : \mathbb{R}^n \to \mathbb{R}$ be polynomial functions with maximal degree d such that the set $[h_1 = 0, \dots, h_r = 0]$ satisfies MFCQ (i.e, the first regularity assumption in Definition 2.1). Then, with a minor adaptation of the above proof, we can show that for the perturbed sets $[g_i \leqslant \alpha, g_j \leqslant 0, i \in I, j \in J] \cap [h_1 = 0, \dots, h_r = 0]$ the number of singular perturbations $\alpha \in \mathbb{R}$ is bounded by

$$d\,(2d-1)^{n+r}(2d+1)^m.$$

Indeed, if α is singular, then there exists a tuple $(x, \lambda, \kappa, \alpha) \in \mathbb{R}^n \times \mathbb{R}^K_{++} \times \mathbb{R}^r \times \mathbb{R}$ that is a solution of a polynomial system similar to (3.5) with the following changes:

- add the r equality constraints $h_j(x) = 0$, $j = 1, \dots, r$;
- replace the right-hand side of the first equality by the linear combination $\sum_{j=1}^r \kappa_j \nabla h_j(x)$.

The rest of the proof follows along the exact same lines, with a trivial adaptation of the notation. We do not need here to take into account the values of the coefficients κ_j, $j = 1, \dots, r$, contrary to those of the λ_i, $i \in K$. Using the same notation as in the proof, this new system has degree d and $n + k_1 + k_2 + r + 1$ variables. Whence the bound follows.

The Milnor–Thom bound (3.4) and a fortiori the bound in Theorem 3.10 are not sharp, but one may ask whether they are of the right order of magnitude. The following examples show that this is indeed the case, at least regarding the dependence with respect to the degree d of the polynomials and the dimension n of the base space. They also illustrate the absence of sensitivity of our bound with respect to the choice of the partition (I, J).

Indeed, the examples show that even if all but one of the constraints define a regular set, the number of singular perturbations generated by the last constraint is of the right order. In the first example, which is thoroughly explained, the degree is fixed to $d = 2$, and the number of singular perturbations is shown to be exponential with respect to n. In the second example, the number of singular diagonal perturbations is shown to be highly dependent on the degree d.

Example 3.12. Here, we construct an inequality constraint set in $\mathbb{R}^n$ defined by $n+1$ polynomial functions of degree 2, n of which are convex. The number of singular perturbations corresponding to a variation of the unique nonconvex constraint is $3^n - 1$.

Let $a \in \mathbb{R}^n$ be a point in $(-1, 1)^n$. Then, for $\alpha \in \mathbb{R}$, define the constraint set $C_{0,\alpha}$ as the set of points $x \in \mathbb{R}^n$ such that

$$\begin{cases} g_0(x) = 4n - \displaystyle\sum_{i=1}^n (x_i - a_i)^2 \leqslant \alpha, \\ g_i(x) = (x_i)^2 \leqslant 1, \quad i = 1, \dots, n. \end{cases}$$

For $\alpha < 4n$, the first inequality defines the complement in $\mathbb{R}^n$ of the open ball centered at point a with radius $\sqrt{4n - \alpha}$, denoted by $B(a, \sqrt{4n - \alpha})$. As for the last n inequalities, they are convex and define the hypercube $[-1, 1]^n$. Observe that for $\alpha \leqslant 0$, $C_{0,\alpha}$ is empty since $[-1, 1]^n$ is strictly included in $B(a, \sqrt{4n - \alpha})$, whereas for $\alpha \geqslant 4n$, $C_{0,\alpha} = [-1, 1]^n$.

[5] Our proof shows that it evolves within the interval $[1/3, 1]$; see (3.6).

We next show that a perturbation α is singular whenever a face of $[-1,1]^n$ and the ball $B(a,\sqrt{4n-\alpha})$ are tangent. First note that the constraint sets $[g_0 \leqslant \alpha]$ and $[g_1 \leqslant 1,\ldots,g_n \leqslant 1]$ both satisfy MFCQ. Hence, the constraint qualification for $C_{0,\alpha}$ may fail only at points where the constraint g_0 and at least one of the constraints g_i with $i \in \{1,\ldots,n\}$ are active, that is, at intersection points between the boundary of the hypercube $[-1,1]^n$ and the boundary of the ball $B(a,\sqrt{4n-\alpha})$.

Let z be such a point for a given α. There exists a nonempty subset of indices I and integers $v_i \in \{\pm 1\}$ for $i \in I$ such that $z_i = v_i$ for all $i \in I$ and $|z_j| < 1$ for all $j \notin I$. Then MFCQ is not satisfied at z if and only if the convex hull of the gradients $\nabla g_0(z)$ and $\nabla g_i(z)$ with $i \in I$ contains 0, and since $[-1,1]^n$ is qualified, this is equivalent to $-\nabla g_0(z)$ being in the convex cone generated by the gradients $\nabla g_i(z)$, $i \in I$. Since $-\nabla g_0(z) = 2(z-a)$ and $\nabla g_i(z) = 2v_i e_i$ for every $i \in I$, where e_i denotes the ith coordinate vector of $\mathbb{R}^n$, the latter condition holds if and only if $z_j = a_j$ for all $j \notin I$, i.e., if and only if z is the orthogonal projection of a on the face $F = \{x \in \mathbb{R}^n \mid x_i = v_i,\ i \in I,\ |x_j| \leqslant 1,\ j \notin I\}$. In other words, MFCQ is not satisfied at point z if and only if a face of the hypercube $[-1,1]^n$ and the ball $B(a,\sqrt{4n-\alpha})$ are tangent at z.

Now, given $k \in \{0,\ldots,n-1\}$ there are $\binom{n}{k}2^{n-k}$ faces of dimension k in the cube $[-1,1]^n$. Thus, by adequately choosing a in $(-1,1)^n$ so that, for all α, $B(a,\sqrt{4n-\alpha})$ is tangent to a unique face of $[-1,1]^n$ at most, we deduce that the total number of singular perturbations is $\sum_{k=0}^{n-1}\binom{n}{k}2^{n-k} = 3^n - 1$. Figure 2 shows a representation of $C_{0,\alpha}$ in $\mathbb{R}^2$ for each singular value α.

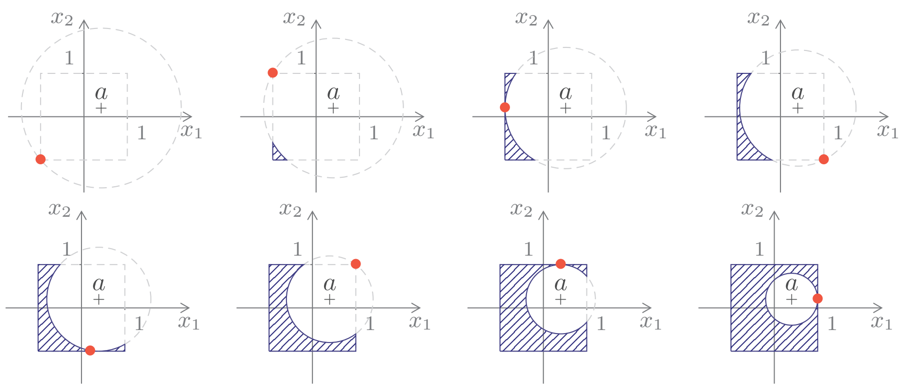

FIG. 2. *Singular perturbations of a constraint set (hatched area) defined by degree* 2 *polynomials.*

Remark 3.13. The dependence of the number of singular values in the previous example, $3^n - 1$, with respect to m and n does not appear clearly since $m = n+1$. In this regard, the gap between this number and the bound predicted by Theorem 3.10, $2 \times 3^n \times 5^{n+1} = 10 \times 15^n$, questions the relevance of the exponential term in m appearing in Theorem 3.10. In order to better understand this dependence, we could think of an example similar to Example 3.12 where the hypercube would be replaced by a polytope with m facets, and hence is defined by m linear constraints (instead of $2n$ in Example 3.12). However, the maximum number of vertices of such a polytope, given by the upper bound theorem [37], is asymptotically equal to $O(m^{\lfloor n/2 \rfloor})$ (see [46]). Hence, such an example could not have a number of singular perturbations exponential with respect to m. It then remains an open question to understand the dependence of the maximum number of singular values with respect to m, n.

Example 3.14. We build an example in $\mathbb{R}^n$ with $n+1$ polynomial constraints, degree $2d$, and we show that the perturbation of a unique constraint generates at least d^n singular values.

For any even integer d, let us consider the polynomial $Q_d = \prod_{k=1}^{d}(X^2 - k^2)$. Let H be the set of points $x \in \mathbb{R}^n$ such that $g_i(x) = Q_d(x_i) \leqslant 0$, $i = 1, \dots, n$. The set H is a disjoint union of d^n boxes. More precisely, since d is even, Q_d is nonpositive on the intervals $[2k-1, 2k]$, $k = 1, \dots, d/2$, and on their symmetrical images with respect to 0. Then H is the (disjoint) union of the d^n boxes

$$H(v,k) := \prod_{i=1}^{n} v_i\,[2k_i - 1, 2k_i], \quad v \in \{\pm 1\}^n, \quad k \in \{1, \dots, d/2\}^n. \tag{3.7}$$

Note that all the boxes (3.7) are included in $[-d, d]^n$. Let a be some point in $(-d, d)^n$ and, for $\alpha \in \mathbb{R}$, define $C_{0,\alpha}$ as the set of points that are contained in H and that satisfy in addition $g_0(x) = 4nd^2 - \|x - a\|^2 \leqslant \alpha$. Figure 3 displays $C_{0,\alpha}$ for $d = 4$.

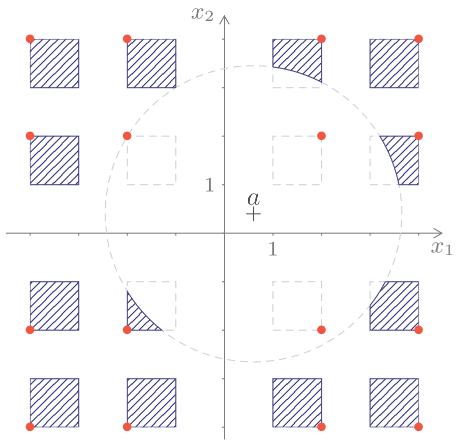

FIG. 3. *Constraint set (hatched area) defined by $n+1$ polynomials of degree at most $2d$ with d^n singular perturbations ($n = 2$, $d = 4$).*

Let us follow the arguments of Example 3.12. Observe that for each vertex $v \in \{\pm 1\}^n$ and for each tuple of indices $k \in \{1, \dots, d/2\}^n$ there exists a unique perturbation $\alpha \in (0, 4nd^2)$ such that MFCQ does not hold at point $(2k_i\, v_i)_{1 \leqslant i \leqslant n}$, that is, when the box $H(v,k)$ defined in (3.7) and the sphere centered at a with radius $\sqrt{4nd^2 - \alpha}$ have a unique contact point (see Figure 3). Finally, by adequately choosing a, it is possible to show that all the d^n singular perturbations mentioned above are distinct.

4. Applications to optimization algorithms. We illustrate here the results of section 3 through some classical algorithms for nonlinear optimization. Our approach consists in embedding the original problem within some one-parameter family of optimization problems:

$$(\mathcal{P}_\alpha) \qquad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g_1(x) \leqslant \alpha, \dots, g_m(x) \leqslant \alpha, \end{array}$$

where $f, g_1, \dots, g_m : \mathbb{R}^m \to \mathbb{R}$ are differentiable definable functions.

A first obvious but important consequence is that any algorithm which is operational under the standard qualification condition can be applied to problem $(\mathcal{P}_\alpha)$ except perhaps for a finite number of parameters α. In view of the fact that $\text{val}(\mathcal{P}_\alpha)$ tends to $\text{val}(\mathcal{P}_0)$ (see Lemma 2.14), it provides a natural way of approximating $(\mathcal{P}_0)$. This can be illustrated in a straightforward manner with many types of algorithms; see, e.g., [6, 30, 42] and see also [2] for continuation techniques in optimization. Estimating the complexity of such an approach is a matter for future research.[6] In this spirit of a direct approximation, we provide an illustration involving SDP relaxations on the KKT ideal, which improves a series of results of [1, 19, 20].

Another family of applications considered below is provided by infeasible SQP methods, which often require strong qualification conditions assumptions.

4.1. Infeasible sequential quadratic programming. We consider the *extended sequential quadratic method*, ESQM, proposed by Auslender [3] and based on an ℓ^∞ penalty function. Other methods could be treated such as, for instance, Flechter's $S\ell^1$QP [26]. We make the following very basic assumptions.

Assumption 4.1.

(i) *Regularity.* The functions $f, g_1, \ldots, g_m : \mathbb{R}^n \to \mathbb{R}$ are $\mathcal{C}^2$ with Lipschitz continuous gradients. We denote by $L, L_1, \ldots, L_m > 0$ their Lipschitz constants, respectively.

(ii) *Compactness.* The constraint sets $C_\alpha = [g_1 \leqslant \alpha, \ldots, g_m \leqslant \alpha]$ are compact and nonempty for all $\alpha \geqslant 0$.

(iii) *Boundedness.* $\inf_{x\in\mathbb{R}^n} f(x) > -\infty$.

The general SQP method we consider, ESQM, is described below. The strength of the following general convergence theorem is to rely merely on semialgebraicity/definability and boundedness assumptions. In particular, it does not require any qualification assumptions whatsoever. Another distinctive feature of this result is to allow us to treat all at once many issues such as nonconvexity, continuum of stationary points, infeasibility, nonlinear constraints, or oscillations (see [12] for more on the key issues).

ESQM Extended sequential quadratic method [3, 12].

Step 1: Choose $x_0 \in C_\alpha$, $\beta_0 > 0$, $\delta > 0$, $\lambda \geqslant L$, and $\lambda' \geqslant \max_i L_i$, and set $k \leftarrow 0$.

Step 2: Compute x_{k+1} solution (along with some $s \in \mathbb{R}$) of

$$\begin{aligned} \underset{s\in\mathbb{R},\, y\in\mathbb{R}^n}{\text{minimize}} \quad & f(x_k) + \langle \nabla f(x_k), y - x_k\rangle + \beta_k s + \frac{\lambda + \beta_k\lambda'}{2}\|y - x_k\|^2 \\ \text{s.t.} \quad & g_i(x_k) + \langle \nabla g_i(x_k), y - x_k\rangle \leqslant \alpha + s\ , \quad i = 1, \ldots, m, \\ & s \geqslant 0. \end{aligned}$$

Step 3: If $g_i(x_k) + \langle \nabla g_i(x_k), x_{k+1} - x_k\rangle \leqslant \alpha$, $i = 1, \ldots, m$, then $\beta_{k+1} \leftarrow \beta_k$. Else $\beta_{k+1} \leftarrow \beta_k + \delta$.

Step 4: $k \leftarrow k + 1$, go to Step 2.

THEOREM 4.2 (large penalty parameters yield convergence of ESQM). *Assume that* (i)–(iii) *hold (smoothness, compactness, boundedness). For all parameters $\alpha \geqslant 0$, except for a finite number of them, there exists a number $\beta(\alpha) \geqslant 0$ such that ESQM*

[6]It is likely to be connected to the results from [27] and [28].

initialized with any $\beta_0 \geqslant \beta(\alpha)$ *generates a sequence* $(x_k)_{k\in\mathbb{N}}$ *which converges to some KKT point of problem* $(\mathcal{P}_\alpha)$.

Proof. Following Corollary 3.7, there exist a finite family of parameters $\mathcal{A} \subset \mathbb{R}_+$ such that for all $\alpha \notin \mathcal{A}$ MFCQ holds throughout C_α. Let us fix a parameter $\alpha \in \mathbb{R}_+ \setminus \mathcal{A}$. Then there exists a positive real number ε such that $[\alpha, \alpha+\varepsilon] \subset \mathbb{R}_+ \setminus \mathcal{A}$. This implies that for every $x \in C_{\alpha+\varepsilon}$, if $g_j(x) \geqslant \alpha$ for some index j, then there exist $y \in \mathbb{R}^n$ such that $\langle y, \nabla g_i(x)\rangle < 0$ for all indices i such that $g_i(x) = \max_{1\leqslant \ell \leqslant m} g_\ell(x)$. This follows from the fact that $x \in C_{\alpha'}$, where $\alpha' = \max_{1\leqslant i\leqslant m} g_i(x)$ is such that $\alpha \leqslant \alpha' \leqslant \alpha+\varepsilon$, so that MFCQ holds at $x \in C_{\alpha'}$.

Let $f_{\min} = \inf_{x\in\mathbb{R}^n} f(x)$ and g_0 be the constant function equal to α. Set $d_k = x_{k+1} - x_k$. For every $k \in \mathbb{N}$, we have

$$\begin{aligned}
\frac{1}{\beta_{k+1}}&(f(x_{k+1}) - f_{\min}) + \max_{0\leqslant i\leqslant m} g_i(x_{k+1}) \\
&\leqslant \frac{1}{\beta_k}(f(x_{k+1}) - f_{\min}) + \max_{0\leqslant i\leqslant m} g_i(x_{k+1}) \\
&\leqslant \frac{1}{\beta_k}(f(x_k) + \langle \nabla f(x_k), d_k\rangle - f_{\min}) \\
&\qquad\qquad + \max_{0\leqslant i\leqslant m} \big(g_i(x_k) + \langle \nabla g_i(x_k), d_k\rangle\big) + \frac{\lambda + \beta_k\lambda'}{2\beta_k}\|d_k\|^2 \\
&\leqslant \frac{1}{\beta_k}(f(x_k) - f_{\min}) + \max_{0\leqslant i\leqslant m} g_i(x_k),
\end{aligned}$$

where the second inequality comes from the Lipschitz continuity of the gradients of the functions involved and the third inequality follows from the minimization problem in Step 2 of ESQM.

Now choose $\beta_0 \geqslant (f(x_0) - f_{\min})/\varepsilon$. By a trivial induction, we deduce that, for every integer $k \in \mathbb{N}$,

$$\max_{0\leqslant i\leqslant m} g_i(x_k) \leqslant \frac{1}{\beta_0}(f(x_0) - f_{\min}) + \max_{0\leqslant i\leqslant m} g_i(x_0) \leqslant \alpha + \varepsilon.$$

Hence, all the points x_k generated by ESQM with the latter choice of β_0 lie in $C_{\alpha+\varepsilon}$ and so satisfy the following qualification condition (an essential ingredient in [12]): if $g_j(x) \geqslant \alpha$ for some index j and some x in $\mathbb{R}^n$, then there exists $y \in \mathbb{R}^n$ such that $\langle y, \nabla g_i(x)\rangle < 0$ for all indices i such that $g_i(x) = \max_{1\leqslant \ell\leqslant m} g_\ell(x)$.

The fact that any cluster point of $(x_k)_{k\in\mathbb{N}}$ is a KKT point of problem $(\mathcal{P}_\alpha)$ readily follows from [12, Thm. 2] (see also [3, Thm. 3.1]). The convergence of $(x_k)_{k\in\mathbb{N}}$ then follows from [12, Thm. 3] and the definability assumptions. □

Remark 4.3 (stabilization of penalty parameters).

(a) For a fixed α, the sequence of penalty parameters β_k is constant after a finite number of iterations. This was an essential result in [3] which still holds here.
(b) As in [12], rates of convergence are available when the data are in addition real semialgebraic.

4.2. Exact relaxation in polynomial programming. A standard approach for solving problem $(\mathcal{P}_\alpha)$ when data are polynomial relies on hierarchies of semidefinite programming (see [33, 34]). It is known that, generically, these hierarchies are exact, meaning that they converge in a finite number of steps (see [41]), but this behavior cannot be detected a priori. In order to construct SDP hierarchies with guaranteed finite convergence behavior, some authors introduced redundant constraints in the hierarchies. The work presented in [20] investigates unconstrained problems and the convergence of SDP hierarchies over the variety of critical points, while [19] considers more generally KKT ideals. The recent work [1] extends these results further and proposes a relaxation which is either exact or which detects in finitely many steps the absence of "KKT minimizers" [1, Thm. 6.3].

A drawback of this method is that it fails whenever optimal solutions of problem $(\mathcal{P}_\alpha)$ do not satisfy KKT conditions.[7] Corollary 3.7 shows that this issue is only a concern for finitely many values of the perturbation parameter α in $(\mathcal{P}_\alpha)$ and that the relaxation remains exact outside of this finite set. We point out that the constructions presented in [19, 20], similar in their approach, require much stronger assumptions on the constraint ideal than the one we propose.

We now explain these facts; Appendix B contains the basic notation/definition used below. We first describe the polynomial problem from which the relaxation in [1] is constructed. Let $\alpha \in \mathbb{R}$ be such that $C_\alpha = [g_1 \leqslant \alpha, \ldots, g_m \leqslant \alpha]$ is nonempty. The Lagrangian associated with problem $(\mathcal{P}_\alpha)$ is defined for $x \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}^m$ by

$$L^\alpha(x, \lambda) := f(x) + \sum_{i=1}^{m} \lambda_i \,(g_i(x) - \alpha).$$

Then we introduce the KKT ideal defined on $\mathbb{R}[x, \lambda]$ by

$$I^\alpha_{\text{KKT}} := \left\langle \frac{\partial L^\alpha}{\partial x_1}, \ldots, \frac{\partial L^\alpha}{\partial x_n}, \lambda_1\,(g_1 - \alpha), \ldots, \lambda_m\,(g_m - \alpha) \right\rangle.$$

Let $\{h^\alpha_1, \ldots, h^\alpha_r\} \subset \mathbb{R}[x]$ be a generating family of the ideal $I^\alpha_{\text{KKT}} \cap \mathbb{R}[x]$:

$$\langle h^\alpha_1, \ldots, h^\alpha_r \rangle = I^\alpha_{\text{KKT}} \cap \mathbb{R}[x].$$

Note that such a family can be obtained by computing a Gröbner basis of I^α_{KKT} (see [17]). Adding these redundant constraints to problem $(\mathcal{P}_\alpha)$ yields the following polynomial problem:

$$(\mathcal{P}^{\text{KKT}}_\alpha) \qquad \begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && g_1(x) \leqslant \alpha, \ldots, g_m(x) \leqslant \alpha, \\ &&& h^\alpha_1(x) = 0, \ldots, h^\alpha_r(x) = 0. \end{aligned}$$

Observe that any minimizer of problem $(\mathcal{P}_\alpha)$ that is also a KKT point is a minimizer of problem $(\mathcal{P}^{\text{KKT}}_\alpha)$. Hence, if the Mangasarian–Fromovitz constraint qualification holds throughout C_α, then solving the former problem boils down to solving the latter.

We next introduce the SDP relaxation hierarchies proposed in [1] to solve problem $(\mathcal{P}^{\text{KKT}}_\alpha)$. For $k \in \mathbb{N}$, the primal is given by

$$(4.1) \quad \begin{aligned} p^\alpha_k = \inf \big\{ \Lambda(f) \mid \Lambda \in (\mathbb{R}_{2k}[x])^*,\ \Lambda(1) = 1,& \\ \Lambda(p) \geqslant 0 \ \ \forall p \in \langle h^\alpha_1, \ldots, h^\alpha_r \rangle_{2k} + \mathfrak{P}_k(\alpha - g_1, \ldots, \alpha - g_m) \big\}& \end{aligned}$$

and the dual problem is

$$(4.2) \qquad d^\alpha_k = \sup \big\{ \gamma \in \mathbb{R} \mid f - \gamma \in \langle h^\alpha_1, \ldots, h^\alpha_r \rangle_{2k} + \mathfrak{P}_k(\alpha - g_1, \ldots, \alpha - g_m) \big\},$$

[7] Abril Bucero and Mourrain gave hints to deal with such a situation, but at the expense of an increasing complexity in the construction of the hierarchies.

where the notation for the truncated ideal $\langle\cdot\rangle_{2k}$ and the truncated preordering $\mathfrak{P}_k$ is detailed in Appendix B. Let us mention however that the dual relaxation hierarchy is based on an SOS representation of nonnegative polynomials, which uses a Schmüdgen-type certificate. But contrary to Schmüdgen's Positivstellensatz [44, Cor. 3], compactness is not required here.

A straightforward combination of Corollary 3.7 and [1, Thm. 6.3] leads to the following.

PROPOSITION 4.4. *Let $f, g_1, \dots, g_m : \mathbb{R}^n \to \mathbb{R}$ be polynomial functions such that $C_0 = [g_1 \leqslant 0, \dots, g_m \leqslant 0]$ is nonempty. Then, for all parameters $\alpha \geqslant 0$, except for a finite number of them, one of the following assertions holds:*

(i) *the relaxations* (4.1) *and* (4.2) *of problem* $(\mathcal{P}_\alpha^{\text{KKT}})$ *are exact and provide the value of problem* $(\mathcal{P}_\alpha)$, *i.e.,* $\text{val}(\mathcal{P}_\alpha) = d_k^\alpha = p_k^\alpha$ *for all k large enough;*[8]

(ii) *for k large enough, the feasible set of problem* (4.1) *is empty and problem* $(\mathcal{P}_\alpha)$ *has no minimizer.*

Appendix A. Reminder on semialgebraic and tame geometry. We recall here the basic results of tame geometry that we use in the present work. Some references on this topic are [16, 23, 24].

DEFINITION A.1 (see [16, Def. 1.4]). *An* o-minimal *structure on* $(\mathbb{R}, +, \cdot)$ *is a sequence of Boolean algebras $\mathcal{O} = (\mathcal{O}_p)_{p\in\mathbb{N}}$ where each $\mathcal{O}_p$ is a family of subsets of $\mathbb{R}^p$ and such that, for each $p \in \mathbb{N}$, we have the following:*

(i) *if A belongs to $\mathcal{O}_p$, then $A \times \mathbb{R}$ and $\mathbb{R} \times A$ belong to $\mathcal{O}_{p+1}$;*

(ii) *if $\pi : \mathbb{R}^{p+1} \to \mathbb{R}^p$ is the canonical projection onto $\mathbb{R}^p$, then, for any $A \in \mathcal{O}_{p+1}$, the set $\pi(A)$ belongs to $\mathcal{O}_p$;*

(iii) *$\mathcal{O}_p$ contains the family of real algebraic subsets of $\mathbb{R}^p$, that is, every set of the form $\{x \in \mathbb{R}^p \mid g(x) = 0\}$, where $g : \mathbb{R}^p \to \mathbb{R}$ is a polynomial function;*

(iv) *the elements of $\mathcal{O}_1$ are exactly the finite unions of points and intervals.*

A subset of $\mathbb{R}^p$ which belongs to an o-minimal structure $\mathcal{O}$ is said to be *definable* (in $\mathcal{O}$). A function $f : A \subset \mathbb{R}^p \to \mathbb{R}^q$ or a set-valued mapping $F : A \subset \mathbb{R}^p \rightrightarrows \mathbb{R}^q$ is said to be definable in $\mathcal{O}$ if its graph is definable (in $\mathcal{O}$) as a subset of $\mathbb{R}^p \times \mathbb{R}^q$.

Example A.2. The simplest (and smallest) o-minimal structure is given by the class $\mathcal{SA}$ of real *semialgebraic* objects. A set $A \subset \mathbb{R}^p$ is called semialgebraic if it is of the form $A = \bigcup_{j=1}^{l} \bigcap_{i=1}^{k} \{x \in \mathbb{R}^p \mid g_{ij}(x) < 0,\ h_{ij}(x) = 0\}$, where the functions $g_{ij}, h_{ij} : \mathbb{R}^p \to \mathbb{R}$ are polynomial functions. The fact that $\mathcal{SA}$ is an o-minimal structure relies mainly on the Tarski–Seidenberg principle (see [5]), which asserts that (ii) holds true in this class.

Other examples like globally subanalytic sets or sets belonging to the log-exp structure provide a vast field of sets and functions that are of primary importance for optimizers. We will not give proper definitions of these structures in this paper, but the interested reader may consult [24] or [8, 9, 32] for optimization-oriented subjects.

In this paper, we shall essentially use the classical results listed hereafter. In the remainder of this subsection, we fix an o-minimal structure $\mathcal{O}$ on $(\mathbb{R}, +, \cdot)$.

PROPOSITION A.3 (stability results). *Let $A \subset \mathbb{R}^p$ and $g : A \to \mathbb{R}^p$ be definable objects.*

[8]The result of Abril Bucero and Mourrain is actually more precise and establishes a link between the minimizers of problem (4.1) and the ones of problem $(\mathcal{P}_\alpha)$. We refer the reader to [1, Thm. 6.3] for a comprehensive presentation.

– *If* $B \subset A$ *is a definable set, then* $g(B)$ *is definable.*
– *If* $C \subset \mathbb{R}^q$ *is a definable set, then* $g^{-1}(C)$ *is definable.*
– *If* A *is open and* g *is differentiable, then its derivative is definable.*

MONOTONICITY LEMMA A.4. *Let* $f : I \subset \mathbb{R} \to \mathbb{R}$ *be a definable function and let* $k \in \mathbb{N}$. *Then there exists a finite partition of* I *into* p *disjoint intervals* $I_1, \dots, I_p$ *such that the restriction of* f *to each nontrivial interval* I_j, $j \in \{1, \dots, p\}$, *is* $\mathcal{C}^k$ *and either constant or strictly monotone.*

DEFINABLE CHOICE LEMMA A.5. *Let* $A \subset \mathbb{R}^p \times \mathbb{R}^q$ *be a definable set and let* $\pi : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}^p$ *be the canonical projection onto* $\mathbb{R}^p$. *Then there exists a definable function* $f : \pi(A) \to \mathbb{R}^q$ *such that* $\operatorname{graph} f \subset A$.

Note that an equivalent formulation of the latter result can be stated in terms of selection: if $F : \mathbb{R}^p \rightrightarrows \mathbb{R}^q$ is a definable set-valued mapping, then there exists a definable function $f : \operatorname{dom} F \to \mathbb{R}^q$ such that $\operatorname{graph} f \subset \operatorname{graph} F$.

CURVE SELECTION LEMMA A.6. *Let* $A \subset \mathbb{R}^p$ *be a definable set, let* x *be an element of* $\operatorname{cl}(A)$, *the topological closure of* A, *and let* $k \in \mathbb{N}$ *be a fixed integer. Then there exists a* $\mathcal{C}^k$ *definable path* $\gamma : [0, 1) \to \mathbb{R}^p$ *such that* $\gamma(0) = x$ *and* $\gamma((0, 1)) \subset A$.

Appendix B. Relaxation in polynomial programming: Definitions and notation. By $\mathbb{R}[x]$ we denote the ring of real polynomials in the variable $x = (x_1, \dots, x_n)$. For any $k \in \mathbb{N}$, we denote by $\mathbb{R}_k[x]$ the space of real polynomials whose degree is bounded by k and we denote by $(\mathbb{R}_k[x])^*$ its dual space.

A polynomial $p \in \mathbb{R}[x]$ is a sum of squares (SOS) if p can be written as $p = \sum_{i \in I} p_i^2$ for some finite family of polynomials $(p_i)_{i \in I} \subset \mathbb{R}[x]$. Denote by $\Sigma[x]$ the space of SOS polynomials.

Given any integer $k \in \mathbb{N}$ and any finite family $\{p_1, \dots, p_m\} \subset \mathbb{R}[x]$ of polynomials, the *k-truncated ideal* on $\mathbb{R}[x]$ generated by this family is the subset of $\mathbb{R}[x]$ defined by

$$\langle p_1, \dots, p_r \rangle_k := \left\{ \sum_{i=1}^{m} q_i\, p_i \;\middle|\; q_i \in \mathbb{R}[x],\ \deg(q_i\, p_i) \leqslant k,\ i = 1, \dots, m \right\},$$

where $\deg(p)$ denotes the degree of any polynomial $p \in \mathbb{R}[x]$. The ideal generated by the family $\{p_1, \dots, p_m\}$ is denoted and defined in a similar way but with no condition required on the degree of the polynomials.

For a set $I \subset \{1, \dots, m\}$, we denote by $p_I \in \mathbb{R}[x]$ the polynomial defined by $p_I := \prod_{i \in I} p_i$, with the convention that $p_\emptyset = 1$. Then we define the *k-truncated preordering* of $\{p_1, \dots, p_m\}$ by

$$\mathfrak{P}_k(p_1, \dots, p_m) := \left\{ \sum_{I} q_I\, p_I \;\middle|\; q_I \in \Sigma[x],\ \deg(q_I\, p_I) \leqslant 2k,\ \forall I \subset \{1, \dots, m\} \right\}.$$

Acknowledgments. The authors thanks J. Pang, H. Frankowska, T. S. Pham, and J.-B. Lasserre for their useful comments.

REFERENCES

[1] M. ABRIL BUCERO AND B. MOURRAIN, *Exact Relaxation for Polynomial Optimization on Semi-Algebraic Sets*, preprint, arXiv.1307.6426, 2014.
[2] E. L. ALLGOWER AND K. GEORG, *Numerical Continuation Methods: An Introduction*, Springer Ser. Comput. Math. 13, Springer, Berlin, 1990, https://doi.org/10.1007/978-3-642-61257-2.

[3] A. Auslender, *An extended sequential quadratically constrained quadratic programming algorithm for nonlinear, semidefinite, and second-order cone programming*, J. Optim. Theory Appl., 156 (2013), pp. 183–212, https://doi.org/10.1007/s10957-012-0145-z.

[4] R. Benedetti, F. Loeser, and J.-J. Risler, *Bounding the number of connected components of a real algebraic set*, Discrete Comput. Geom., 6 (1991), pp. 191–209, https://doi.org/10.1007/BF02574685.

[5] R. Benedetti and J.-J. Risler, *Real Algebraic and Semi-Algebraic Sets*, Actualités Math., Hermann, Paris, 1990.

[6] D. P. Bertsekas, *Nonlinear Programming*, Athena Sci. Optim. Comput. Ser., 3rd ed., Athena Scientific, Belmont, MA, 2016.

[7] L. Blum, M. Shub, and S. Smale, *On a theory of computation and complexity over the real numbers: NP-completeness, recursive functions and universal machines*, Bull. Amer. Math. Soc. (N.S.), 21 (1989), pp. 1–46, https://doi.org/10.1090/S0273-0979-1989-15750-9.

[8] J. Bolte, A. Daniilidis, and A. Lewis, *The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems*, SIAM J. Optim., 17 (2007), pp. 1205–1223, https://doi.org/10.1137/050644641.

[9] J. Bolte, A. Daniilidis, and A. Lewis, *Tame functions are semismooth*, Math. Program., 117 (2009), pp. 5–19, https://doi.org/10.1007/s10107-007-0166-9.

[10] J. Bolte, A. Daniilidis, A. Lewis, and M. Shiota, *Clarke subgradients of stratifiable functions*, SIAM J. Optim., 18 (2007), pp. 556–572, https://doi.org/10.1137/060670080.

[11] J. Bolte, A. Daniilidis, and A. S. Lewis, *Generic optimality conditions for semialgebraic convex programs*, Math. Oper. Res., 36 (2011), pp. 55–70, https://doi.org/10.1287/moor.1110.0481.

[12] J. Bolte and E. Pauwels, *Majorization-minimization procedures and convergence of SQP methods for semi-algebraic and tame programs*, Math. Oper. Res., 41 (2016), pp. 442–465, https://doi.org/10.1287/moor.2015.0735.

[13] J.-F. Bonnans, J. C. Gilbert, C. Lemaréchal, and C. A. Sagastizábal, *Numerical Optimization: Theoretical and Practical Aspects*, Universitext, 2nd ed., Springer, Berlin, 2006.

[14] J.-F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer Ser. Oper. Res., Springer, New York, 2000.

[15] J. M. Borwein and A. S. Lewis, *Convex Analysis and Nonlinear Optimization*, CMS Books Math./Ouvrages de Math. SMC 3, 2nd ed., Springer, New York, 2006, .

[16] M. Coste, *An Introduction to O-Minimal Geometry*, RAAG notes, Institut de Recherche Mathématique de Rennes, 1999; available at https://perso.univ-rennes1.fr/michel.coste/polyens/OMIN.pdf.

[17] D. A. Cox, J. Little, and D. O'Shea, *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*, Undergrad. Texts Math., 4th ed., Springer, New York, 2015, https://doi.org/10.1007/978-3-319-16721-3.

[18] A. Daniilidis and J. C. H. Pang, *Continuity and differentiability of set-valued maps revisited in the light of tame geometry*, J. Lond. Math. Soc., 83 (2011), pp. 637–658.

[19] J. Demmel, J. Nie, and V. Powers, *Representations of positive polynomials on noncompact semialgebraic sets via KKT ideals*, J. Pure Appl. Algebra, 209 (2007), pp. 189–200, https://doi.org/10.1016/j.jpaa.2006.05.028.

[20] J. Demmel, J. Nie, and B. Sturmfels, *Minimizing polynomials via sum of squares over the gradient ideal*, Math. Program., 106 (2006), pp. 587–606, https://doi.org/10.1007/s10107-005-0672-6.

[21] A. L. Dontchev, M. Quincampoix, and N. Zlateva, *Aubin criterion for metric regularity*, J. Convex Anal., 13 (2006), pp. 281–297.

[22] A. L. Dontchev and R. T. Rockafellar, *Implicit Functions and Solution Mappings*, Springer Ser. Oper. Res. Financ. Eng., 2nd ed., Springer, New York, 2014, https://doi.org/10.1007/978-1-4939-1037-3.

[23] L. van den Dries, *Tame Topology and O-Minimal Structures*, London Math. Soc. Lecture Note Ser. 248, Cambridge University Press, Cambridge, 1998, https://doi.org/10.1017/CBO9780511525919.

[24] L. van den Dries and C. Miller, *Geometric categories and o-minimal structures*, Duke Math. J., 84 (1996), pp. 497–540, https://doi.org/10.1215/S0012-7094-96-08416-1.

[25] D. Drusvyatskiy, A. D. Ioffe, and A. S. Lewis, *Generic minimizing behavior in semialgebraic optimization*, SIAM J. Optim., 26 (2016), pp. 513–534, https://doi.org/10.1137/15M1020770.

[26] R. Fletcher, *An ℓ_1 penalty method for nonlinear constraints*, in Numerical Optimization 1984, SIAM, Philadelphia, PA, 1985, pp. 26–40.

[27] H. Frankowska, *Some inverse mapping theorems*, Ann. Inst. H. Poincaré Anal. Non Linéaire, 7 (1990), pp. 183–234.

[28] H. Frankowska and M. Quincampoix, *Hölder metric regularity of set-valued maps*, Math. Program., 132 (2012), pp. 333–354, https://doi.org/10.1007/s10107-010-0401-7.

[29] O. Fujiwara, *Morse programs: A topological approach to smooth constrained optimization.* I, Math. Oper. Res., 7 (1982), pp. 602–616.

[30] P. E. Gill and E. Wong, *Sequential Quadratic Programming Methods*, Springer, New York, 2012, pp. 147–224, https://doi.org/10.1007/978-1-4614-1927-3_6.

[31] H. V. Hà and T. S. Pham, *Genericity in Polynomial Optimization*, Ser. Optim. Appl. 3, World Scientific, Hackensack, NJ, 2017.

[32] A. Ioffe, *A Sard theorem for tame set-valued mappings*, J. Math. Anal. Appl., 335 (2007), pp. 882–901, https://doi.org/10.1016/j.jmaa.2007.01.104.

[33] J. B. Lasserre, *Global optimization with polynomials and the problem of moments*, SIAM J. Optim., 11 (2001), pp. 796–817, https://doi.org/10.1137/S1052623400366802.

[34] J. B. Lasserre, *Moments, Positive Polynomials and Their Applications*, Imp. Coll. Press Optim. Ser. 1, Imperial College Press, London, 2010.

[35] G. M. Lee and T.-S. Pham, *Generic properties for semialgebraic programs*, Optim. Online, 2015; available at http://www.optimization-online.org/DB_FILE/2015/06/4957.pdf.

[36] G. M. Lee and T.-S. Pham, *Stability and genericity for semi-algebraic compact programs*, J. Optim. Theory Appl., 169 (2016), pp. 473–495, https://doi.org/10.1007/s10957-016-0910-5.

[37] P. McMullen, *The maximum numbers of faces of a convex polytope*, Mathematika, 17 (1970), pp. 179–184, https://doi.org/10.1112/S0025579300002850.

[38] J. W. Milnor, *Topology from the Differentiable Viewpoint*, University Press of Virginia, Charlottesville, VA, 1965.

[39] B. S. Mordukhovich, *Maximum principle in the problem of time optimal response with nonsmooth constraints*, Prikl. Mat. Meh., 40 (1976), pp. 1014–1023.

[40] B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation* I, Grundlehren Math. Wiss. 330, Springer, Berlin, 2006.

[41] J. Nie, *Optimality conditions and finite convergence of Lasserre's hierarchy*, Math. Program., 146 (2014), pp. 97–121, https://doi.org/10.1007/s10107-013-0680-x.

[42] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer Ser. Oper. Res. Financ. Eng., 2nd ed., Springer, New York, 2006.

[43] R. T. Rockafellar and R. J.-B. Wets, *Variational analysis*, Grundlehren Math. Wiss. 317, Springer, Berlin, 1998, https://doi.org/10.1007/978-3-642-02431-3.

[44] K. Schmüdgen, *The K-moment problem for compact semi-algebraic sets*, Math. Ann., 289 (1991), pp. 203–206, https://doi.org/10.1007/BF01446568.

[45] S. Scholtes and M. Stöhr, *How stringent is the linear independence assumption for mathematical programs with complementarity constraints?*, Math. Oper. Res., 26 (2001), pp. 851–863, https://doi.org/10.1287/moor.26.4.851.10007.

[46] R. Seidel, *The upper bound theorem for polytopes: An easy proof of its asymptotic version*, Comput. Geom., 5 (1995), pp. 115–116, https://doi.org/10.1016/0925-7721(95)00013-Y.

[47] J. E. Spingarn and R. T. Rockafellar, *The generic nature of optimality conditions in nonlinear programming*, Math. Oper. Res., 4 (1979), pp. 425–430, https://doi.org/10.1287/moor.4.4.425.