

# SENSITIVITY ANALYSIS FOR PARAMETRIC NONLINEAR PROGRAMMING: A TUTORIAL

FRANÇOIS PACAUD \*

**Abstract.** This tutorial provides an overview of the current state-of-the-art in the sensitivity analysis for nonlinear programming. Building upon the fundamental work of Fiacco, it derives the sensitivity of primal-dual solutions for regular nonlinear programs and explores the extent to which Fiacco's framework can be extended to degenerate nonlinear programs with non-unique dual solutions. The survey ends with a discussion on how to adapt the sensitivity analysis for conic programs and approximate solutions obtained from interior-point algorithms.

**1. Introduction.** Sensitivity analysis has been a long standing research thread in nonlinear programming. It aims at evaluating the sensitivity of a solution  $x \in \mathbb{R}^n$  of a nonlinear problem with respect to parameter changes. For a parameter  $p \in \mathbb{R}^\ell$ , the solution of a generic optimization problem is given as solution of a (possibly nonsmooth) variational problem:

$$(1.1) \quad F(x, p) = 0 ,$$

with  $F : \mathbb{R}^n \times \mathbb{R}^\ell \rightarrow \mathbb{R}^n$  a function encoding the stationary conditions of the optimization problem (e.g. the Karush-Kuhn-Tucker stationary conditions, or the homogeneous self-dual embedding associated to a conic program). The solution of (1.1) depends on  $p$  and is noted  $x(p)$ . If the function  $F$  is *regular* enough, the Implicit Function Theorem gives an elegant way to evaluate the Jacobian  $J_p x(p) \in \mathbb{R}^{n \times \ell}$  of the solution  $x(p)$  of (1.1) as the solution of the linear system:

$$(1.2) \quad J_p x(p) = -(J_x F(x, p))^{-1} J_p F(x, p) .$$

Most applications require only the forward and adjoint sensitivities, alleviating the need to compute the full Jacobian  $J_p x(p)$ . For vectors  $u \in \mathbb{R}^\ell$  and  $v \in \mathbb{R}^n$ , the forward and adjoint sensitivities are defined respectively as

$$(1.3) \quad \begin{aligned} J_p x(p) u &= -(J_x F(x, p))^{-1} \bar{u} \quad \text{with } \bar{u} := J_p F(x, p) u , \\ J_p x(p)^\top v &= -J_p F(x, p)^\top \bar{v} \quad \text{with } \bar{v} := (J_x F(x, p))^{-\top} v . \end{aligned}$$

Following [8], we call *implicit differentiation* the class of differentiation methods based on the relation (1.2). Implicit differentiation requires solving the nonlinear system (1.1) exactly and then relies on the Implicit Function Theorem: it is accurate, but often limited in scope. A practical alternative uses the fact that the system (1.1) is usually solved *iteratively*, e.g. using an optimization solver. If the iterative method is differentiable (in the sense each of its operation can be differentiated by automatic differentiation), the sensitivity  $J_p x(p)$  can be computed *iteratively* using forward accumulation (also known as *piggy back* method [31]). This second mode does not rely on the Implicit Function Theorem, as it propagates the derivatives directly through the algorithm. The main advantage is the derivatives are directly computed at the solution, meaning we do not have to solve an additional linear system as in (1.2). However, the method can be sensitive to numerical error and the derivatives  $J_p x(p)$  converge more slowly than the primal solution  $x(p)$  [29, 30].

---

\*Centre Automatique et Systèmes (CAS), Mines Paris-PSL.

Hence, implicit differentiation remains the favorite method to compute the sensitivities. Unfortunately, the hypothesis of the Implicit Function Theorem are often too stringent to hold for a generic nonlinear program, limiting the applicability of (1.2) unless practical workarounds are applied. For nonlinear programs, the problem's regularity is related to its non-degeneracy: if we perturb slightly the parameters, we have to ensure that the primal solution exists and remains unique. Depending on the structure of the problem, the perturbed solution may have the following demonstrable properties [24], ordered by difficulty:

- Existence of the primal solution.
- Existence and continuity of the primal solution.
- (Upper) Lipschitz continuity of the primal solution, non-uniqueness of the primal-dual solution.
- (Upper) Lipschitz continuity of the primal-dual solution, non-uniqueness of the primal-dual solution.
- Directional differentiability and uniqueness of the primal solution, non-uniqueness of the dual solution.

In this tutorial, we aim at summarizing the conditions under which the sensitivity analysis is well-posed.

### 1.1. A brief history of sensitivity analysis for nonlinear programming.

We start by a brief review of literature to remind the main developments in the theory of sensitivity analysis for nonlinear programming.

*The 1970s: defining regularity and differentiability.* The regularity of the KKT solution has been studied extensively in the 1970s by Robinson [45, 46], by interpreting the nonlinear program as a generalized nonlinear equation. It is now established that the regularity of a nonlinear program usually follows from the constraint qualification conditions satisfied at the KKT solution. Robinson proved in [46] that the primal-dual solution is *strongly regular* (hence, locally unique and Lipschitz continuous) if LICQ and SSOSC holds at the original solution. In addition, the (Fréchet) differentiability of the primal-dual solution is established if SCS hold, as established by Fiacco in his seminal result [19].

*The 1980s and 1990s: tackling degeneracy.* The previous results have been extended in the 1980s, with a primary focus on dropping the strong SCS condition required in [19]. Kojima [36] proved that MFCQ and GSSOSC are sufficient to guarantee that the primal-dual solution is strongly stable. Moreover, the primal-dual solution becomes non-differentiable as soon as we drop the SCS condition, but it remains *directionally differentiable* under certain assumptions. Jittorntrum [35] showed that if LICQ and SSOSC hold, then the directional derivative of the primal-dual solution exists and is given as the solution of a quadratic program (QP).

If we drop the LICQ condition, the dual solution becomes non-unique and the directional differentiability of the dual solution is lost. Shapiro [48] showed that the primal solution remains directionally differentiable if we replace LICQ by SMCQ and SSOSC. This result was extended by Kyparisis in [39], who proved that we retain directionally differentiability for the primal solution if we assume MFCQ, CRCQ and GSSOSC, that is, without requiring unicity of the dual solution. This thread of research culminated in 1995 with the publication of the seminal result of Ralph and Dempe [43], giving a practical method to evaluate the directional derivative of the primal solution under MFCQ.

*The 2000s and 2010s: extending the scope of applications.* The theory of sensitivity analysis for nonlinear program has witnessed a significant consolidation after

the 1990s, with the publication of three consequent books dedicated to the sensitivity analysis of optimization programs [10, 18, 17]. In particular, the application of sensitivity analysis has been extended to the more general setting offered by variational inequalities, which covers both nonlinear programs and complementarity problems. In parallel, numerous applications have flourished out of the sensitivity analysis theory, one of the most significant being bilevel programming [16].

**1.2. Motivations.** Sensitivity analysis has proven to be a fertile ground for the analysis of stochastic optimization problems [49, 50], parametric decomposition [27, 13], parameter estimation problems [40] and last but not least, model predictive control [58, 34, 57]. Recently, the field has witnessed a renewed interest with the emergence of the *differentiable programming* paradigm in the machine learning community, where people look at embedding a (convex) optimization solver inside a neural network [3, 1] (whose training requires the evaluation of the sensitivity in the backward pass).

**1.3. Outline.** Section 2 is devoted to the theory of nonlinear programming, with an emphasis on the constraint qualifications used to define the regularity of the feasible set. Section 3 summarizes the main results about the sensitivity of the primal-dual solution in nonlinear programming. It extends Fiacco's seminal work [21] and puts a special consideration on degenerate nonlinear programs, following the lead of Ralph and Dempe [43]. Section 4 focuses more particularly on the sensitivity of the objective function. Last, Section 5 details practical methods and specific software implementations to compute the sensitivity of any optimization program, including an extended discussion on the sensitivity of conic programs.

**2. Nonlinear programming.** This section recalls classical results in the theory of nonlinear programming, with a long discussion on constraint qualifications §2.2. The scope of this tutorial is limited to finite-dimensional optimization problems. We refer to [10] for a broader description covering the infinite dimensional cases.

**2.1. Notations.** For a parameter  $p \in \mathbb{R}^\ell$ , we are interested in solving the parametric nonlinear optimization problem:

$$(2.1) \quad \text{NLP}(p) := \min_{x \in \mathbb{R}^n} f(x, p) \quad \text{subject to} \quad \begin{cases} g(x, p) = 0, \\ h(x, p) \leq 0, \end{cases}$$

where  $f : \mathbb{R}^n \times \mathbb{R}^\ell \rightarrow \mathbb{R}$  an objective and  $g : \mathbb{R}^n \times \mathbb{R}^\ell \rightarrow \mathbb{R}^{m_e}$  and  $h : \mathbb{R}^n \times \mathbb{R}^\ell \rightarrow \mathbb{R}^{m_i}$  two functions encoding respectively the equality and inequality constraints.

All functions are twice-differentiable and depend jointly on the optimization variable  $x \in \mathbb{R}^n$  and a parameter  $p \in \mathbb{R}^\ell$ . The *feasible set* of the optimization problem (2.1) is encoded by the set-valued function

$$(2.2) \quad X(p) = \{x \in \mathbb{R}^n \mid g(x, p) = 0, h(x, p) \leq 0\}.$$

*Value function.* We define the value function  $\varphi : \mathbb{R}^\ell \rightarrow \mathbb{R} \cup \{+\infty\}$  as the solution of NLP( $p$ ) for a given parameter  $p \in \mathbb{R}^\ell$ :

$$(2.3) \quad \varphi(p) = \min_{x \in \mathbb{R}^n} f(x, p) \quad \text{subject to} \quad x \in X(p).$$

By convention  $\varphi(p) = +\infty$  if the problem is infeasible ( $X(p) = \emptyset$ ). The *solution set* associated to (2.3) is noted

$$(2.4) \quad \Sigma(p) := \{x \in X(p) : f(x, p) = \varphi(p)\}.$$

*Lagrangian.* We introduce the Lagrangian associated to (2.1) as

$$(2.5) \quad \mathcal{L}(x, y, z, p) = f(x, p) + y^\top g(x, p) + z^\top h(x, p),$$

where  $y \in \mathbb{R}^{m_e}$  (resp.  $z \in \mathbb{R}^{m_i}$ ) the multiplier associated to the equality (resp. inequality) constraints. The Karush-Kuhn-Tucker (KKT) conditions of (2.1) state that under a certain constraint qualification, the primal vector  $x$  is a local optimal solution if there exists a dual multiplier  $(y, z)$  such that

$$(2.6) \quad \begin{aligned} \nabla_x f(x, p) + \sum_{i=1}^{m_e} y_i \nabla_x g_i(x, p) + \sum_{i=1}^{m_i} z_i \nabla_x h_i(x, p) &= 0, \\ g(x, p) &= 0, \quad h(x, p) \leq 0, \quad z \geq 0, \quad z^\top h(x, p) = 0. \end{aligned}$$

The KKT conditions become necessary and sufficient in the convex case. We define a primal-dual stationary solution (or KKT stationary solution) as the vector

$$(2.7) \quad w(p) = (x(p), y(p), z(p)) \in \mathbb{R}^n \times \mathbb{R}^{m_e} \times \mathbb{R}^{m_i}.$$

To alleviate the notations, we often denote by  $w^*$  the vector  $w(p)$ .

*Multiplier set.* The set of multipliers  $(y, z) \in \mathbb{R}^{m_e+m_i}$  satisfying the KKT conditions (2.6) is denoted by  $\mathcal{M}_p(x)$ . We notice that the set  $\mathcal{M}_p(x)$  is polyhedral (hence closed and convex). We note by  $\mathcal{E}_p(x)$  the set of extreme points of the polyhedron  $\mathcal{M}_p(x)$ . We say that the problem is *degenerate* if  $\mathcal{M}_p(x)$  has more than one element.

*Active set.* For a given  $x \in \mathbb{R}^n$ , we define the *active set* at  $x$  as

$$(2.8) \quad \mathcal{I}_p(x) = \{i \in [m_i] \mid h_i(x, p) = 0\}.$$

For given multipliers  $(y, z) \in \mathcal{M}_p(x)$ , we define the *strongly* and *weakly active constraints* as

$$(2.9) \quad \mathcal{I}_p^+(x; z) = \{i \in \mathcal{I}_p(x) \mid z_i > 0\}, \quad \mathcal{I}_p^0(x; z) = \{i \in \mathcal{I}_p(x) \mid z_i = 0\}.$$

The strongly and weakly active constraints form a partition of the active set:  $\mathcal{I}_p(x) = \mathcal{I}_p^+(x; z) \cup \mathcal{I}_p^0(x; z)$  and  $\mathcal{I}_p^+(x; z) \cap \mathcal{I}_p^0(x; z) = \emptyset$ .

Degeneracy arises from the undetermined weakly active constraints  $\mathcal{I}_p^0(x; z)$ : a small perturbation of the parameter  $p$  is likely to render them inactive. We say that the problem satisfies *strict complementarity slackness* (SCS) if there is no weakly active constraint:  $\mathcal{I}_p^0(x; z) = \emptyset$ .

*Critical cone.* The critical cone is a key notion to characterize the second-order optimality conditions of a nonlinear program.

**DEFINITION 2.1** (Critical cone). *For a feasible point  $x \in X(p)$ , the critical cone associated to the feasible set  $X(p)$  at  $x$  is defined as*

$$(2.10) \quad \begin{aligned} \mathcal{C}_p(x) = \Big\{ d \in \mathbb{R}^n : & \nabla_x f(x, p)^\top d \leq 0, \\ & \nabla_x g_i(x, p)^\top d = 0, \forall i \in [m_e], \\ & \nabla_x h_j(x, p)^\top d \leq 0, \forall j \in \mathcal{I}_p(x) \Big\}. \end{aligned}$$

We note that at a KKT solution  $(x^*, y^*, z^*)$ , the critical cone simplifies as

$$(2.11) \quad \begin{aligned} \mathcal{C}_p(x^*) = \Big\{ d \in \mathbb{R}^n : & \nabla_x g_i(x^*, p)^\top d = 0, \forall i \in [m_e], \\ & \nabla_x h_j(x^*, p)^\top d = 0, \forall j \in \mathcal{I}_p^+(x^*; z^*), \\ & \nabla_x h_j(x^*, p)^\top d \leq 0, \forall j \in \mathcal{I}_p^0(x^*; z^*) \Big\}. \end{aligned}$$

Further, if we assume that strict complementarity slackness holds, the critical cone is exactly the null space of the active constraints' Jacobian:  $\mathcal{C}_p(x^*) = N(J_{act}(x^*))$ :

$$(2.12) \quad \begin{aligned} \mathcal{C}_p(x^*) = \{d \in \mathbb{R}^n : \nabla_x g_i(x^*, p)^\top d = 0, \forall i \in [m_e], \\ \nabla_x h_j(x^*, p)^\top d = 0, \forall j \in \mathcal{I}_p(x^*)\} . \end{aligned}$$

In parametric optimization, it is common to introduce the critical set associated to the critical cone (2.10) at the solution [18, 43].

**DEFINITION 2.2** (Critical set). *Let  $(x^*, y^*, z^*)$  be a KKT stationary solution. The critical set at the solution is the polyhedral cone defined as*

$$(2.13) \quad \begin{aligned} \mathcal{K}_p(x^*, z^*) = \{d \in \mathbb{R}^{n+\ell} : \nabla_{(x,p)} g_i(x^*, p)^\top d = 0, \forall i \in [m_e], \\ \nabla_{(x,p)} h_j(x^*, p)^\top d = 0, \forall j \in \mathcal{I}_p^+(x^*; z^*), \\ \nabla_{(x,p)} h_j(x^*, p)^\top d \leq 0, \forall j \in \mathcal{I}_p^0(x^*; z^*)\} . \end{aligned}$$

*Elements of  $\mathcal{K}_p(x^*, z^*)$  are called critical directions. The directional critical set in the direction  $h \in \mathbb{R}^\ell$  is defined as*

$$(2.14) \quad \mathcal{K}_p(x^*, z^*; h) = \{d \in \mathbb{R}^n : (d, h) \in \mathcal{K}_p(x^*, z^*)\} .$$

**2.2. Constraint qualifications.** Constraint qualifications ensure that we can capture *locally* the geometry of the feasible set  $X(p)$  by linearizing it near a feasible point  $x \in X(p)$ . The regularity of a nonlinear program is defined by the particular constraint qualifications that hold at a stationary point. We recall in this section the main constraint qualifications we can encounter in practice, and their respective impacts on the problem's topology.

**DEFINITION 2.3** (Linear-independence constraint qualification). *LICQ holds if the gradients of active constraints*

$$(2.15) \quad \{\nabla_x g_i(x, p) : i \in [m_e]\} \cup \{\nabla_x h_i(x, p) : i \in \mathcal{I}_p(x)\} ,$$

*are linearly independent.*

LICQ implies the following second-order necessary condition.

**THEOREM 2.4.** *Suppose  $(x^*, y^*, z^*)$  is a primal-dual solution of (2.1) satisfying LICQ. Then,*

$$(2.16) \quad d^\top \nabla_{xx}^2 L(x, y, z, p) d \geq 0 \quad \forall d \in \mathcal{C}_p(x) ,$$

*where  $\mathcal{C}_p(x)$  is the critical cone of the feasible set  $X(p)$  at  $x$ .*

**DEFINITION 2.5** (Mangasarian-Fromovitz constraint qualification). *MFCQ holds if the gradients of the equality constraints*

$$(2.17) \quad \{\nabla_x g_i(x, p) : i \in [m_e]\} ,$$

*are linearly independent and if there exists a vector  $d \in \mathbb{R}^{n_x}$  such that*

$$(2.18) \quad \nabla_x g_i(x, p)^\top d = 0, \forall i \in [m_e], \quad \nabla_x h_i(x, p)^\top d < 0, \forall i \in \mathcal{I}_p(x) .$$

It is well known that MFCQ generalizes the Slater's condition of convex optimization to nonlinear programming. MFCQ also implies the boundedness of the optimal multipliers set.

THEOREM 2.6 ([26]). *The set  $\mathcal{M}_p(x)$  is bounded if and only if MFCQ holds.*

PROPOSITION 2.7 (Dual MFCQ). *Using the theorem of alternatives, MFCQ is equivalent to 0 being the unique solution of the linear system*

$$(2.19) \quad \begin{aligned} \sum_{i=1}^{m_e} y_i \nabla_x g_i(x, p) + \sum_{i \in \mathcal{I}_p(x)} z_i \nabla_x h_i(x, p) &= 0, \\ z_i &\geq 0 \quad \forall i \in \mathcal{I}_p(x). \end{aligned}$$

DEFINITION 2.8 (Strict Mangasarian-Fromovitz constraint qualification). *SMFCQ holds if*

$$(2.20) \quad \{\nabla_x g_i(x, p) : i \in [m_e]\} \cup \{\nabla_x h_i(x, p) : i \in \mathcal{I}_p^+(x; z)\},$$

*are linearly independent and if there exists a vector  $d \in \mathbb{R}^{n_x}$  such that*

$$(2.21) \quad \begin{aligned} \nabla_x g_i(x, p)^\top d &= 0, \forall i \in [m_e], \\ \nabla_x h_i(x, p)^\top d &= 0, \forall i \in \mathcal{I}_p^+(x; z), \\ \nabla_x h_i(x, p)^\top d &< 0, \forall i \in \mathcal{I}_p^0(x; z). \end{aligned}$$

A primal-dual solution  $(x^*, y^*, z^*)$  satisfying SMFCQ implies a second-order necessary condition analogous to Theorem 2.4 [38, Theorem 2.1]. SMFCQ is the weakest condition under which the multiplier set  $\mathcal{M}_p(x)$  reduces to a singleton.

PROPOSITION 2.9 (Proposition 1.1, [38]). *Let  $w = (x, y, z)$  be a primal-dual solution. Then SMFCQ holds at  $x$  if and only if the multiplier vector  $(y, z)$  is unique.*

DEFINITION 2.10 (Constant-rank constraint qualification (CRCQ)). *CRCQ holds if there exists a neighborhood  $W$  of  $x^*$  such that for any subsets  $I$  of  $\mathcal{I}_p(x^*)$  and  $J$  of  $[m_e]$  the family of gradient vectors*

$$(2.22) \quad \{\nabla_x g_j(x, p) : j \in J\} \cup \{\nabla_x h_i(x, p) : i \in I\},$$

*has the same rank for all vectors  $x \in W$ .*

We now focus on second-order constraint qualifications, starting with the second-order sufficiency condition.

DEFINITION 2.11 (Second-order sufficiency condition (SOSC)). *SOSC holds at  $x$  if there exists  $(y, z) \in \mathcal{M}_p(x)$  such that*

$$(2.23) \quad d^\top \nabla_{xx}^2 L(x, y, z, p) d > 0 \quad \forall d \in \mathcal{C}_p(x), d \neq 0.$$

PROPOSITION 2.12. *Suppose SOSC holds at a solution  $x^*$ . Then, there exists  $\sigma > 0$  and a neighborhood  $V$  of  $x^*$  such that for all  $x \in X(p) \cap V$  (with  $x \neq x^*$ ),*

$$(2.24) \quad f(x) > f(x^*) + \frac{\sigma}{2} \|x - x^*\|^2.$$

We define the set

$$(2.25) \quad \mathcal{D}_p(x; z) = \{d \in \mathbb{R}^{n_x} : \nabla_x g_i(x, p) d = 0 \quad \forall i \in [m_e], \\ \nabla_x h_j(x, p) d = 0 \quad \forall j \in \mathcal{I}_p^+(x; z)\}.$$

Note that at a primal-dual solution  $(x, y, z)$ ,  $\mathcal{C}_p(x; z) \subset \mathcal{D}_p(x; z)$  with both sets being equal if SCS holds.

DEFINITION 2.13 (Strong second-order sufficiency condition (SSOSC)). *SSOSC holds at  $x$  if there exists  $(y, z) \in \mathcal{M}_p(x)$  such that*

$$(2.26) \quad d^\top \nabla_{xx}^2 L(x, y, z, p) d > 0 \quad \forall d \in \mathcal{D}_p(x; z), d \neq 0.$$

*If the condition (2.26) hold for all multipliers in  $\mathcal{M}_p(x)$ , then we say that the generalized second-order sufficiency condition (GSSOSC) hold.*

If SCS hold, we note that SSOSC becomes equivalent to SOSC.

**3. Sensitivity of the primal-dual solution.** Let  $p^* \in \mathbb{R}^\ell$  be a parameter, and  $w^* = (x^*, y^*, z^*)$  a KKT stationary solution of NLP( $p$ ). We aim at studying the behavior of the primal-dual solution  $w(p) := (x(p), y(p), z(p))$  near  $p^*$ . In particular, we are interested in characterizing the local continuity of the solution  $w(p)$  (i.e., *is  $w(p)$  Lipschitz-continuous?*) as well as its local differentiability (i.e., *under which conditions is  $w(p)$  differentiable?*).

We start by recalling Fiacco's theorem in §3.1. This result is often restrictive, as it does not allow for active set changes in the solution. For that reason, the sensitivity are often derived using Robinson's generalized equations, as described in §3.2. The generalized equation framework is a convenient way to encode the nonsmoothness in the problem's solution. In §3.3, we show that it can also be interpreted using the notion of lexicographic derivative. The last section §3.4 is devoted to the differentiation of degenerate nonlinear program, where the dual solution is not unique. We show that in that case the notion of  $PC^1$  functions is the most adapted to characterize the nature of the solution.

**3.1. Fiacco and the implicit function theorem.** If the problem is regular, the local continuity is usually given by the Implicit Function Theorem.

THEOREM 3.1 (Implicit function theorem). *Let  $F : \mathbb{R}^n \times \mathbb{R}^\ell \rightarrow \mathbb{R}^n$  be a  $C^1$  function and  $(x^*, p^*) \in \mathbb{R}^n \times \mathbb{R}^\ell$  such that  $F(x^*, p^*) = 0$ . If the Jacobian  $J_x F(x^*, p^*)$  is invertible, then there exists a neighborhood  $U \subset \mathbb{R}^\ell$  of  $p^*$  and a differentiable function  $x(p)$  such that  $F(x(p), p) = 0$  for all  $p \in U$ , and  $x(p)$  is the unique solution in a neighborhood of  $x^*$ . In addition,  $x(\cdot)$  is  $C^1$  differentiable on  $U$ , with*

$$(3.1) \quad J_p x(p) = -(J_x F(x, p))^{-1} J_p F(x, p).$$

In [19], Fiacco investigated under which conditions the KKT equations falls under the conditions of the implicit function theorem. To do so, we should ensure (i) the primal solution mapping  $x(p)$  is single-valued (i.e. the problem (2.1) has an unique solution) (ii) the dual solution  $(y(p), z(p))$  is unique and (iii) the active set is locally stable. Fiacco showed that these three conditions are satisfied if respectively (i) SOSC (ii) LICQ and (iii) SCS hold.

In that particular case, the KKT equations (2.6) rewrite as a smooth system of nonlinear equations depending continuously on  $w = (x, y, z)$ . For  $i \in \mathcal{I}_p(x)$  we have  $z_i > 0$  and  $h_i(x, p) = 0$ ; alternatively for  $i \in [m_i] \setminus \mathcal{I}_p(x)$  we have  $z_i = 0$  and  $h_i(x, p) < 0$ . By differentiating the complementarity condition  $z_i h_i(x, p) = 0$ , we get for all  $i = 1, \dots, m_i$ ,

$$(3.2) \quad \frac{\partial z_i}{\partial p} \cdot h_i(x, p) + z_i \cdot \left( \frac{\partial h_i}{\partial x} \frac{\partial x}{\partial p} + \frac{\partial h_i}{\partial p} \right) = 0.$$

Hence, if  $i \in [m_i] \setminus \mathcal{I}_p(x)$  we get by elimination  $\frac{\partial z_i}{\partial p} = 0$ : the multiplier  $z_i = 0$  remains constant after a small perturbation, meaning the constraints remain inactive.

Alternatively, if  $i \in \mathcal{I}_p(x)$ , we deduce that  $\frac{\partial h_i}{\partial x} \frac{\partial x}{\partial p} + \frac{\partial h_i}{\partial p} = 0$  : the constraint remains locally active. As a consequence, we can discard the inactive constraints information from the KKT conditions and interpret the remaining active constraints as equality constraints. We note  $z_{\mathcal{I}} = \{z_i\}_{i \in \mathcal{I}_p}$  the vector storing the active multipliers and  $h_{\mathcal{I}}(x, p) = \{h_i(x, p)\}_{i \in \mathcal{I}_p}$  the vector storing the active inequality constraints.

Locally, the solution  $w := (x, y, z_{\mathcal{I}})$  satisfies the smooth nonlinear system:

$$(3.3) \quad F(w) = \begin{pmatrix} \nabla_x \mathcal{L}(x, y, z_{\mathcal{I}}) \\ g(x, p) \\ h_{\mathcal{I}}(x, p) \end{pmatrix} = 0.$$

We note the Jacobian of  $F(\cdot)$  :

$$(3.4) \quad M := J_w F = \begin{bmatrix} \nabla_{xx}^2 \mathcal{L} & J_x g^\top & J_x h_{\mathcal{I}}^\top \\ J_x g & 0 & 0 \\ J_x h_{\mathcal{I}} & 0 & 0 \end{bmatrix}, \quad N := J_p F = \begin{bmatrix} \nabla_{xp}^2 \mathcal{L} \\ J_p g \\ J_p h_{\mathcal{I}} \end{bmatrix}.$$

The sensitivity of the solution is given by the following theorem, which applies the Implicit Function Theorem 3.1 to the system (3.3).

**THEOREM 3.2** (Theorem 3.2.2 [19]). *Suppose that at a KKT stationary solution  $x^*$  of  $NLP(p^*)$ , SOSC, LICQ and SCS hold. Then,*

1. *the primal-dual solution  $w^* = (x^*, y^*, z_{\mathcal{I}}^*)$  is unique;*
2. *there exists a unique continuously differentiable function  $w(\cdot)$  defined in a neighborhood  $U$  of  $p^*$  such that  $w(p^*) = (x^*, y^*, z_{\mathcal{I}}^*)$ , with*

$$(3.5) \quad J_p w(p) = -M^{-1} N.$$

3. *SCS and LICQ hold locally near  $p^*$ .*

In detail, Equation (3.5) gives the sensitivity of the primal-dual solution as

$$(3.6) \quad \begin{bmatrix} J_p x \\ J_p y \\ J_p z_{\mathcal{I}} \end{bmatrix} = - \begin{bmatrix} \nabla_{xx}^2 \mathcal{L} & J_x g^\top & J_x h_{\mathcal{I}}^\top \\ J_x g & 0 & 0 \\ J_x h_{\mathcal{I}} & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} \nabla_{xp}^2 \mathcal{L} \\ J_p g \\ J_p h_{\mathcal{I}} \end{bmatrix}.$$

The operation translates as the solution of a (sparse) linear system with  $\ell$  right-hand-sides. We note that the linear system (3.6) has a saddle point structure: the SOSC and LICQ conditions implies that it is invertible [7, Theorem 3.4]. If the original problem  $NLP(p)$  is solved with a Newton algorithm, the factorization of the Jacobian matrix  $J_w F$  is already available, leaving only to compute  $\ell$  backsolves.

*Remark 3.3* (Linear programming). If the problem (2.1) is a linear program (LP), we have  $\nabla_{xx}^2 \mathcal{L} = 0$ . The Goldman-Tucker theorem states there always exists a LP solution satisfying strict-complementarity. If the LP is not degenerate, LICQ holds and the solution is a vertex of the polyhedral feasibility set  $X(p)$ . This automatically implies SOCP (the critical cone reduces to  $\mathcal{C}_p(x^*) = \{0\}$ ) and there are exactly  $n$  binding constraints:  $m_e + |\mathcal{I}_p(x^*)| = n$ . Hence Theorem 3.2 applies to the linear programming case. We denote the active Jacobian at the solution  $B = [J_x g^\top \quad J_x h_{\mathcal{I}}^\top]^\top$ . The matrix  $B$  has dimension  $n \times n$  and is non-singular (LICQ implies it is full row-rank). We note that  $B$  is exactly the basis matrix used in the simplex algorithm. In that case, the inverse of the matrix  $M = \begin{bmatrix} 0 & B^\top \\ B & 0 \end{bmatrix}$  in (3.5) satisfies:

$$(3.7) \quad M^{-1} = \begin{bmatrix} 0 & B^{-1} \\ B^{-\top} & 0 \end{bmatrix}.$$



Hence, the computation of the sensitivities simplifies considerably if the problem is regular linear program. If the linear program is solved using the simplex algorithm, the solver usually returns a LU factorization for the basis matrix  $B$ .

*Remark 3.4* (Exponential decay). Certain optimization problems have a graph structure: in that case the primal-dual variable can be reordered following a set of nodes  $\mathcal{V}$  such that  $w = \{w_i\}_{i \in \mathcal{V}}$ . As a consequence, the Jacobian matrix  $M$  appearing in (3.5) can also be reordered following the graph structure: in the language of linear algebra, there exists a permutation matrix  $P$  such that the matrix  $P^\top M P$  is *banded*. It is known that terms in the inverse of a banded matrix decay exponentially fast away from the diagonal [14]. Then, assuming SSOSC, SCS and LICQ the solution of (3.5) satisfies the *exponential decay of sensitivity* property [53], that is, for two parameters  $p, p'$  close enough,

$$(3.8) \quad \|w_i(p) - w_i(p')\| \leq \sum_{j \in \mathcal{V}} C_{ij} \|p_j - p'_j\|, \quad \forall i \in \mathcal{V},$$

with  $w_i(p)$  the primal-dual nodal solution at node  $i \in \mathcal{V}$  and  $C_{ij} = \Gamma \rho^{d_G(i,j)}$  (with  $\Gamma > 0$  and  $\rho > 0$  two positive constants) and  $d_G(i, j)$  the graph distance between node  $i$  and node  $j$ . The Equation (3.8) has important consequence for computing the sensitivities in a decentralized fashion [56] or building a preconditionner to solve (3.5) with an iterative method [52].

**3.2. Robinson's generalized equations.** We emphasize that the strict complementarity condition is unlikely to hold on practical optimization problems, where constraints can change locally from active to inactive. For that reason, it has been investigated how to generalize Fiacco's result without the strict complementarity assumption. In [44], Robinson proposed a powerful framework to characterize the solution of nonlinear program (2.1), by interpreting the KKT conditions (2.6) as a *generalized equation*  $F(w, p) + N_C(w) \ni 0$  (see Appendix C for a brief recall on generalized equations).

**PROPOSITION 3.5** (KKT system as generalized equation). *Let  $w = (x, y, z) \in \mathbb{R}^n \times \mathbb{R}^{m_e} \times \mathbb{R}^{m_i}$  be a primal-dual variable. A primal-dual solution  $w$  of the KKT system (2.6) is also solution of the generalized equation*

$$(3.9) \quad F(w, p) + N_C(w) \ni 0,$$

$$\text{with } F(w, p) := \begin{bmatrix} \nabla_x \mathcal{L}(x, y, z, p) \\ -\nabla_y \mathcal{L}(x, y, z, p) \\ -\nabla_z \mathcal{L}(x, y, z, p) \end{bmatrix} \text{ and } C = \mathbb{R}^n \times \mathbb{R}^{m_e} \times \mathbb{R}_+^{m_i}.$$

When writing the KKT system as a generalized equation, the nonsmoothness arising from the active set changes is encoded implicitly by the normal cone  $N_C(x)$ ,  $F$  being here a smooth functional. If the generalized equation (3.9) is strongly regular, then the sensitivities follow from Robinson's Theorem (Theorem C.4 in the appendix). Strong regularity is implied by SSOSC and LICQ, allowing to drop the SCS condition (however, SSOSC is necessary to guarantee that the perturbed solution  $x(p)$  remains unique without SCS).

**PROPOSITION 3.6** (Theorem 4.1, [45]). *Let  $w^* = (x^*, y^*, z^*)$  a primal-dual solution of the generalized equation (3.9). If SSOSC and LICQ hold at  $w^*$ , then (3.9) is strongly regular at  $w^*$ .*

The strong regularity given by Proposition 3.6 implies

- (a) The solution  $x^*$  is an isolated local minimum of  $\text{NLP}(p)$  and the associated Lagrange multipliers  $(y^*, z^*)$  are unique;
- (b) For  $p$  near  $p^*$ , there exists a unique Lipschitz continuous function  $w(p) = (x(p), y(p), z(p))$  satisfying KKT and SOSC for  $\text{NLP}(p)$ , with  $w(p^*) = w^*$ , and  $x(p)$  being a unique solution of  $\text{NLP}(p)$ ;
- (c) LICQ holds at  $x(p)$  for  $p$  near  $p^*$ .

Hence, strict complementarity slackness is not required to ensure the primal-dual solution  $w(p)$  is locally Lipschitz continuous. In addition, we can prove that the optimal solution is also directionally differentiable in a direction  $h \in \mathbb{R}^\ell$ .

**PROPOSITION 3.7** (Theorem 2, [35]). *Suppose that at a KKT stationary solution  $x^*$  of  $\text{NLP}(p^*)$ , LICQ and SSOSC hold. Then*

- *there exists a local unique continuous primal-dual solution  $w(\cdot)$  which is directionally differentiable in every direction  $h \in \mathbb{R}^\ell$ ;*
- *for a given direction  $h \in \mathbb{R}^\ell$ , the directional derivative  $w'(p; h)$  is the primal-dual solution of the QP problem defined as*

$$(3.10) \quad \begin{aligned} & \min_d \frac{1}{2} d^\top (\nabla_{xx}^2 \mathcal{L}) d + h^\top (\nabla_{xp}^2 \mathcal{L}) d \\ & \text{s.t.} \quad \nabla_x g_i(x, p)^\top d + \nabla_p g_i(x, p)^\top h = 0 \quad (\forall i \in [m_e]), \\ & \quad h_j(x, p) + \nabla_x h_j(x, p)^\top d + \nabla_p h_j(x, p)^\top h \leq 0 \quad (\forall j \in [m_i]), \end{aligned}$$

$$\text{with } \nabla_{xx}^2 \mathcal{L} = \nabla_{xx}^2 \mathcal{L}(x, y, z, p), \nabla_{xp}^2 \mathcal{L} = \nabla_{xp}^2 \mathcal{L}(x, y, z, p).$$

The QP is parameterized implicitly by the dual multipliers  $(y, z)$  (appearing in the derivatives of the Lagrangian  $\nabla_{xx}^2 \mathcal{L}$  and  $\nabla_{xp}^2 \mathcal{L}$ ). LICQ guarantees that the multipliers are unique at the solution, leading to a non-ambiguous definition of the directional derivative (3.10). Under LICQ and SSOSC, the directional derivative exists for every direction  $h \in \mathbb{R}^\ell$ . However, we cannot guarantee that the directional derivative  $w'(p; \cdot)$  is continuous unless SCS hold. Note that under SCS, the QP (3.10) rewrites with only equality constraints and becomes equivalent to the linear system (3.5).

**3.3. Lexicographic derivative as an alternative to Robinson's generalized equation.** Robinson's generalized equation is a powerful framework to encode nonsmooth equations. Several alternatives have been developed in the recent years, one of the most promising method being based on the notion of lexicographic derivative [41], which has been introduced has a practical differentiation method for nonsmooth functions, such as  $PC^1$  functions (see Appendix A.4). The KKT conditions (2.6) rewrites equivalently as a system of nonsmooth equations:

$$(3.11) \quad \Phi(x, y, z, p) := \begin{bmatrix} \nabla_x L(x, y, z, p) \\ g(x, p) \\ \min\{-h(x, p), z\} \end{bmatrix} = 0.$$

It is well known that the nonsmooth KKT system (3.11) can be solved using a semi-smooth Newton method. At the solution, the sensitivity analysis of the nonsmooth system can be carried out using lexicographic derivatives [6].

The function  $\min$  is  $PC^1$  in the sense of Scholtes [47], so we deduce that the mapping  $\Phi$  is itself  $PC^1$  on its domain. The notion of coherently oriented structure extends the Implicit Function Theorem 3.1 to  $PC^1$  functions.

**PROPOSITION 3.8** (Theorem 5.1 [55]). *Let  $(x^*, y^*, z^*)$  be a solution of KKT. If  $\Phi$  is coherently oriented with respect to  $(x, y, z)$  at  $(x^*, y^*, z^*)$  then there exists a*

neighborhood  $U$  of  $p$  and  $V$  of  $x^*$  and a  $PC^1$  mapping  $w : U \rightarrow V$  such that for each  $p \in U$   $w(p)$  is the unique solution of (3.11). Moreover, for any  $k \in \mathbb{N}$  and any  $R \in \mathbb{R}^{p \times k}$  the LD-derivatives  $x'(p; R)$ ,  $y'(p, R)$  and  $z'(p, R)$  are the unique solution  $X$ ,  $Y$  and  $Z = [Z^+ \ Z^0 \ Z^-]$  of the following nonsmooth equation system

$$(3.12) \quad \begin{bmatrix} \nabla_{xx}^2 \mathcal{L} & G_x^\top & (H_x^+)^{\top} \\ G_x & 0 & 0 \\ H_x^+ & 0 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z^+ \end{bmatrix} = - \begin{bmatrix} \nabla_{xp}^2 \mathcal{L} \\ G_p \\ H_p^+ \end{bmatrix} R, \\ \mathbf{LMmin}(-H_p^0 R - H_x^0 X, Z^0) = 0, \\ Z^- = 0,$$

where  $\mathbf{LMmin}$  is the lexicographic matrix minimum function defined in (A.24), and

$$G_{(x,p)} = J_{(x,p)} g(x, p), \quad H_{(x,p)}^+ = J_{(x,p)} h_{\mathcal{I}_p^+}(x, p), \quad H_{(x,p)}^0 = J_{(x,p)} h_{\mathcal{I}_p^0}(x, p).$$

Without surprise, ensuring complete coherently orientation is equivalent to assuming LICQ and SSOSC hold.

**PROPOSITION 3.9.** *Let  $(x^*, y^*, z^*)$  be a KKT stationary point. Suppose that LICQ and SSOSC holds at  $(x^*, y^*, z^*)$ . Then  $\Phi$  is coherently oriented with respect to  $(x, y, z)$ .*

Solving the nonsmooth system (3.12) is generally non trivial, and falls back to a nonsmooth Newton method or an active set algorithm (cycling through linear equation system solves parameterized by the current working active set). However, it has been proved in [54] that evaluating the LD derivatives can be rewritten as the solution a hierarchy of  $k$  QP problems  $QP_{(1)}, QP_{(2)}, \dots, QP_{(k)}$ , with  $QP_{(j)}(r_{(j)})$  defined for  $j = 1, \dots, k$  as

$$\begin{aligned} \min_{d \in \mathbb{R}^n} \quad & \frac{1}{2} d^\top \nabla_{xx}^2 \mathcal{L} d + r_{(j)}^\top \nabla_{px}^2 \mathcal{L} d \\ \text{subject to} \quad & \nabla_x g_i(x, p)^\top d + \nabla_p g_i(x, p)^\top r_{(j)} = 0 \quad \forall i \in [m_e] \quad \leftarrow \text{multiplier } \alpha \\ & \nabla_x h_i(x, p)^\top d + \nabla_p h_i(x, p)^\top r_{(j)} = 0 \quad \forall i \in \mathcal{A}_{(j-1)}^+ \quad \leftarrow \text{multiplier } \beta \\ & \nabla_x h_i(x, p)^\top d + \nabla_p h_i(x, p)^\top r_{(j)} \leq 0 \quad \forall i \in \mathcal{A}_{(j-1)}^0 \quad \leftarrow \text{multiplier } \gamma \end{aligned}$$

We note  $\alpha_{(j)} \in \mathbb{R}^{m_e}$ ,  $\beta_{(j)} \in \mathbb{R}^{m_i^+}$ ,  $\gamma_{(j)} \in \mathbb{R}^{m_i^0}$  the unique multipliers associated resp. to the equality constraints, the strongly active constraints and the weakly active constraints. The sets of strongly and weakly active constraints change between two consecutive indexes  $j$  using the information brought by the latest multiplier  $\gamma_{(i)}$ :

$$(3.13) \quad \begin{aligned} \mathcal{A}_{(j)}^0 &= \{i \in \mathcal{A}_{(j-1)}^0 : \nabla_x h_i(x, p)^\top d + \nabla_p h_i(x, p)^\top p = 0 \text{ and } \gamma_i = 0\}, \\ \mathcal{A}_{(j)}^+ &= \{i \in \mathcal{A}_{(j-1)}^0 : \nabla_x h_i(x, p)^\top d + \nabla_p h_i(x, p)^\top p = 0 \text{ and } \gamma_i > 0\} \cup \mathcal{A}_{(j-1)}^+, \end{aligned}$$

with  $\mathcal{A}_{(0)}^0 = \mathcal{I}_p^0(x; z)$  and  $\mathcal{A}_{(0)}^+ = \mathcal{I}_p^+(x; z)$ .

**PROPOSITION 3.10** (Theorem 3.1 [54]). *Let  $(x^*, y^*, z^*)$  be a solution of KKT. Suppose that LICQ and SSOSC hold. Then, for any  $k \in \mathbb{N}$  and  $R \in \mathbb{R}^{\ell \times k}$ , the LD-derivatives  $x'(p, R)$ ,  $y'(p, R)$  and  $z'(p, R)$  in the direction  $R$  are given as*

$$(3.14) \quad \begin{aligned} x'(p, R) &= [d_{(1)}(r_{(1)}) \ d_{(2)}(r_{(2)}) \ \dots \ d_{(k)}(r_{(k)})], \\ y'(p, R) &= [\alpha_{(1)}(r_{(1)}) \ \alpha_{(2)}(r_{(2)}) \ \dots \ \alpha_{(k)}(r_{(k)})], \\ z'(p, R) &= [\mu_{(1)}(r_{(1)}) \ \mu_{(2)}(r_{(2)}) \ \dots \ \mu_{(k)}(r_{(k)})], \end{aligned}$$

with, for  $j = 1, \dots, k$ ,

$$(3.15) \quad \begin{aligned} \mu_{(j)}(r_{(j)})_{\mathcal{A}_{(j)}^+} &= \beta_{(j)}(r_{(j)}) , \\ \mu_{(j)}(r_{(j)})_{\mathcal{A}_{(j)}^0} &= \gamma_{(j)}(r_{(j)}) , \\ \mu_{(j)}(r_{(j)})_{\mathcal{A}_{(j)}^-} &= 0 . \end{aligned}$$

**3.4. Towards degeneracy: dropping the LICQ condition.** We have shown that the sensitivities remain well-defined if we drop the SCS condition required in Theorem 3.2. What happens if we suppose in addition that the multipliers  $(y, z)$  are non unique? It turns out that in that case we retain the directional differentiability of the primal solution  $x(p)$ . However, the definition of QP (3.10) becomes ambiguous: which multipliers should we consider when evaluating the Hessian of the Lagrangian?

Kojima was the first to prove the continuity of the optimal solution assuming only MFCQ and GSSOSC, and proved in addition that they are the weakest conditions under which the perturbed solution is locally unique [36].

**THEOREM 3.11** (Theorem 7.2, [36]). *Suppose that at a KKT stationary solution  $x^*$  of  $NLP(p^*)$ , MFCQ and GSSOSC hold. Then, there are open neighborhoods  $U$  of  $p^*$  and  $V$  of  $x^*$  and a function  $x : U \rightarrow V$  such that  $x(\cdot)$  is continuous. The function  $x(p)$  is the unique local solution of  $NLP(p)$  in  $V$ , and MFCQ holds at  $x(p)$ .*

The LICQ condition in Proposition 3.7 can be relaxed to SMFCQ, which also implies the multipliers  $(y^*, z^*)$  at the solution are unique (Proposition 2.9).

**PROPOSITION 3.12** (Theorem 4.2, [48]). *Suppose that at a solution  $x^*$  of  $NLP(p^*)$ , SMFCQ and SSOSC hold. Then,  $x(\cdot)$  is directionally differentiable at  $p$  for every  $h \in \mathbb{R}^\ell$ , and the directional derivative  $x'(p^*, h)$  is the unique solution of the QP problem*

$$(3.16) \quad \min_d \frac{1}{2} d^\top (\nabla_{xx}^2 \mathcal{L}) d + h^\top (\nabla_{xp}^2 \mathcal{L}) d \quad \text{s.t.} \quad d \in \mathcal{K}_{p^*}(x^*, z^*; h) ,$$

with  $\mathcal{K}_{p^*}$  the critical cone defined in (2.10) and  $\nabla_{xx}^2 \mathcal{L} = \nabla_{xx}^2 \mathcal{L}(x^*, y^*, z^*, p^*)$ ,  $\nabla_{xp}^2 \mathcal{L} = \nabla_{xp}^2 \mathcal{L}(x^*, y^*, z^*, p^*)$ .

Note that we are no longer able to evaluate the directional derivative of the dual solution in (3.16), in contrast with the QP (3.10). Kyparisis [39] proved that we can relax SMFCQ by CRCQ to get the existence of the directional derivative  $x'(p, h)$  in the degenerate case where the multipliers  $(y, z)$  are non unique: it suffices to look at the multipliers at the extreme points  $\mathcal{E}_{p^*}(x^*)$ .

**PROPOSITION 3.13** (Theorem 2.2, [39]). *Suppose that at a KKT stationary solution  $x^*$  of  $NLP(p^*)$ , MFCQ, CRCQ and GSSOSC hold. Then,  $x(\cdot)$  is directionally differentiable at  $p^*$  for every  $h \in \mathbb{R}^\ell$ , and the directional derivative  $x'(p^*, h)$  is the unique solution of the QP problem, defined for a given  $(y, z) \in \mathcal{E}_{p^*}(x^*)$ ,*

$$(3.17) \quad \min_d \frac{1}{2} d^\top (\nabla_{xx}^2 \mathcal{L}) d + h^\top (\nabla_{xp}^2 \mathcal{L}) d \quad \text{s.t.} \quad d \in \mathcal{K}_{p^*}(x^*, z; h) ,$$

with  $\mathcal{K}_{p^*}$  the critical cone defined in (2.10) and  $\nabla_{xx}^2 \mathcal{L} = \nabla_{xx}^2 \mathcal{L}(x^*, y, z, p^*)$ ,  $\nabla_{xp}^2 \mathcal{L} = \nabla_{xp}^2 \mathcal{L}(x^*, y, z, p^*)$ .

Finally, Proposition 3.13 has been extended in [43] to derive a practical way to evaluate the sensitivities for a degenerate nonlinear program. The idea is to look only

at multipliers in  $\mathcal{E}_{p^*}(x^*)$  that are also solution of the linear program

$$(3.18) \quad S_{p^*}(x^*; h) := \begin{cases} \arg \max_{y, z} y^\top (J_p g(x^*, p^*)) h + z^\top (J_p h(x^*, p^*)) h \\ \text{s.t. } (y, z) \in \mathcal{M}_{p^*}(x^*) . \end{cases}$$

The solution  $(y, z)$  of (3.18) determines the sets of strongly active  $\mathcal{I}_p^+(x^*; z)$  and weakly active constraints  $\mathcal{I}_p^0(x^*; z)$  in the QP (3.17). As such, the set  $S_{p^*}(x^*; h)$  is intimately related to the critical set  $\mathcal{K}_p(x^*, z; h)$ . Indeed, observe that by definition  $(y, z) \in \mathcal{M}_p(x)$  if

$$(3.19) \quad \nabla_x f(x^*, p^*) + \sum_{i=1}^{m_e} y_i \nabla_x g_i(x^*, p^*) + \sum_{i \in \mathcal{I}_p(x^*)} z_i \nabla_x h_i(x^*, p^*) = 0, \quad z_{\mathcal{I}} \geq 0,$$

where we have removed the inactive inequality from (3.19). By noting  $d \in \mathbb{R}^n$  the multiplier associated to the stationary condition, and  $\lambda \in \mathbb{R}^{|\mathcal{I}_p(x^*)|}$  the multiplier associated to the constraints  $z_{\mathcal{I}} \geq 0$ , the KKT conditions of (3.18) write

$$\begin{aligned} \nabla_x g_i(x^*, p^*)^\top d + \nabla_p g_i(x^*, p^*)^\top h &= 0 & \forall i \in [m_e], \\ \nabla_x h_i(x^*, p^*)^\top d + \nabla_p h_i(x^*, p^*)^\top h + \lambda &= 0 & \forall i \in \mathcal{I}_p(x^*), \\ \nabla_x f(x^*, p^*) + \sum_{i=1}^{m_e} y_i \nabla_x g_i(x^*, p^*) + \sum_{i \in \mathcal{I}_p(x^*)} z_i \nabla_x h_i(x^*, p^*) &= 0, \\ 0 \leq \lambda \perp z_{\mathcal{I}} \geq 0. \end{aligned}$$

By removing the multiplier  $\lambda$  and taking into account the disjunction in the complementarity constraint  $0 \leq \lambda \perp z_{\mathcal{I}} \geq 0$ , the two first equations rewrite equivalently as

$$(3.20) \quad \begin{aligned} \nabla_x g_i(x^*, p^*)^\top d + \nabla_p g_i(x^*, p^*)^\top h &= 0 & \forall i \in [m_e], \\ \nabla_x h_i(x^*, p^*)^\top d + \nabla_p h_i(x^*, p^*)^\top h &\leq 0 & \forall i \in \mathcal{I}_p^0(x^*; z), \\ \nabla_x h_i(x^*, p^*)^\top d + \nabla_p h_i(x^*, p^*)^\top h &= 0 & \forall i \in \mathcal{I}_p^+(x^*; z), \end{aligned}$$

which describes exactly the critical set  $\mathcal{K}_p(x^*, z; h)$ . Hence, by theory of linear programming, the problem (3.18) has an optimal solution if and only if the previous KKT system is consistent, meaning that  $\mathcal{K}_p(x^*, z; h)$  is nonempty. This is formalized in the following lemma.

LEMMA 3.14 (Lemma 2.2, [15]). *The critical cone  $\mathcal{K}_p(x^*, z; h)$  is nonempty if and only if  $(y, z) \in S_{p^*}(x^*; h)$ .*

We can go one step further in the interpretation of the LP (3.18). Note that its dual writes

$$(3.21) \quad \begin{aligned} \min_{d \in \mathbb{R}^n} \quad & \nabla_x f(x^*, p^*)^\top d, \\ \text{s.t.} \quad & \nabla_x g_i(x^*, p^*)^\top d + \nabla_p g_i(x^*, p^*)^\top h = 0, \quad \forall i \in [m_e], \\ & \nabla_x h_i(x^*, p^*)^\top d + \nabla_p h_i(x^*, p^*)^\top h \leq 0, \quad \forall i \in \mathcal{I}_p(x^*), \end{aligned}$$

with the multiplier  $y \in \mathbb{R}^{m_e}$  being associated to the equality constraints and the multiplier  $z \in \mathbb{R}^{m_i}$  being associated to the inequality constraints. As a consequence,

the multipliers  $(y, z)$  are the marginal costs encoding the objective's change after a perturbation  $h$ , and satisfy

$$(3.22) \quad \nabla_x f(x^*, p^*)^\top d + \sum_{i=1}^{m_e} y_i \nabla_p g_i(x^*, p^*)^\top h + \sum_{i \in \mathcal{I}_p(x^*)}^{m_e} z_i \nabla_p h_i(x^*, p^*)^\top h = 0 .$$

The nature of the LP (3.18) has been further studied in [33, Lemma 4.1]. Izmailov and Solodov noted that the solution of (3.18) is likely to be unique if the objective of (3.18) is non-constant. Furthermore, if SMFCQ is not satisfied the set of multipliers  $\mathcal{M}_p(x^*)$  does not reduce to a singleton. Then, the LP (3.18) returns a solution  $(y, z) \in \mathcal{E}_p(x^*)$  that does not satisfy strict-complementarity:  $\mathcal{I}_p^0(x^*, z) \neq \emptyset$ .

Ralph and Dempe showed in [43] that the directional derivative is unique even when the solution of the LP (3.18) is not. They also prove that the local solution  $x(\cdot)$  is piecewise differential in the sense of Scholtes [47].

**THEOREM 3.15** (Theorem 2, [43]). *Suppose that at a KKT stationary solution  $x^*$  of NLP( $p^*$ ), MFCQ, CRCQ and GSSOSC hold. Then,*

- *There is open neighborhoods  $U$  of  $p^*$  and  $V$  of  $x^*$  and a function  $x : U \rightarrow V$  such that  $x(p)$  is the unique local solution of NLP( $p$ ) in  $V$  and MFCQ holds at  $x(p)$ ;*
- *$x(\cdot)$  is a  $PC^1$  function (hence locally Lipschitz and B-differentiable);*
- *For each  $p \in U$ , the directional derivative  $x'(p; \cdot)$  is a piecewise linear function, and for  $h \in \mathbb{R}^\ell$  and  $(y, z) \in S_p(x^*; h)$ ,  $x'(p; h)$  is the unique solution of (3.17).*

It turns out that the notion of  $PC^1$  functions is adapted to characterize the solution of a parametric nonlinear program.

**4. Sensitivity of the optimal value function.** We now focus on the derivative of the value function  $\varphi(p)$ . In §4.1, we recall classical results focusing on the directional differentiability of the value function  $\varphi$ . The differentiable case is studied in §4.2.

**4.1. Danskin lemma and its extensions.** The sensitivity analysis of the optimal value function traces back to the seminal work of Danskin, who studied the sensitivity for a simplified problem where the admissible set  $X(p)$  is a fixed subset  $X \subset \mathbb{R}^n$ , that is,

$$(4.1) \quad \varphi(p) = \min_{x \in X} f(x, p) .$$

**THEOREM 4.1** (Danskin's Lemma [12]). *Suppose  $X$  is nonempty and compact, and  $f(x, \cdot)$  is continuously differentiable at  $p \in \mathbb{R}^\ell$ . Then  $\varphi$  is locally Lipschitz near  $p$ , and directionally differentiable in every direction  $h \in \mathbb{R}^\ell$ , with*

$$(4.2) \quad \varphi'(p; h) = \min_{x \in \Sigma(p)} \nabla_p f(x, p)^\top h ,$$

with  $\Sigma(p)$  the solution set defined in (2.4).

Danskin's theorem has been extended by Hogan to the case where  $X(p)$  is defined as  $X(p) = \{x \in \mathbb{R}^n : x \in M, h(x, p) \leq 0\}$ , for a fixed set  $M$ .

**THEOREM 4.2** ([32]). *Suppose  $M$  is a closed convex set,  $f(\cdot, p)$  and  $h_i(\cdot, p)$  are convex on  $M$  for all  $p$ , and continuously differentiable on  $M \times N$ , with  $N$  a neighborhood of  $p$ . If (i) the solution set  $\Sigma(p)$  is nonempty and bounded, (ii)  $\varphi(p)$  is finite,*

(iii) there is a point  $\hat{x} \in M$  such that  $h(\hat{x}, p) < 0$  (Slater), then  $\varphi$  is directionally differentiable, and for all  $h \in \mathbb{R}^\ell$ ,

$$(4.3) \quad \varphi'(p; h) = \min_{x \in \Sigma(p)} \max_{z \in \mathcal{M}_p(x)} \nabla_p \mathcal{L}(x, z, p)^\top h,$$

with  $\mathcal{M}_p(x)$  the multiplier solution set.

The extension to the generic case associated to  $X(p) = \{x \in \mathbb{R}^n \mid g(x, p) = 0, h(x, p) \leq 0\}$  was proved by Gauvin and Dubeau [28] and Fiacco [20]. The following theorem bounds the lower and upper Dini derivatives (see Appendix A.2) of the value function: in general the bounds are sharp.

**THEOREM 4.3** (Theorem 2.3.4 [21]). *Suppose that  $X(p)$  is nonempty and uniformly compact near  $p$  and MFCQ holds at each  $x \in \Sigma(p)$ . Then  $\varphi$  is locally Lipschitz near  $p$ , and for any  $h \in \mathbb{R}^\ell$ ,*

$$(4.4) \quad \begin{aligned} \inf_{x \in \Sigma(p)} \min_{(y, z) \in \mathcal{M}_p(x)} \nabla_p \mathcal{L}(x, y, z, p)^\top h &\leq \liminf_{t \downarrow 0} \frac{1}{t} (\varphi(p + th) - \varphi(p)) \\ &\leq \limsup_{t \downarrow 0} \frac{1}{t} (\varphi(p + th) - \varphi(p)) \\ &\leq \inf_{x \in \Sigma(p)} \max_{(y, z) \in \mathcal{M}_p(x)} \nabla_p \mathcal{L}(x, y, z, p)^\top h. \end{aligned}$$

In the convex case the multiplier set  $\mathcal{M}_p(x)$  is independent on the solution  $x \in \Sigma(p)$ .

**PROPOSITION 4.4** (Corollary 2.3.8 [21]). *Let  $f$  and  $h$  be convex functions in  $x$ ,  $g$  be affine in  $x$ , such that  $f, g, h$  are all jointly  $C^1$  in  $(x, p)$ . Then, if  $X(p)$  is nonempty and uniformly compact near  $p$  and MFCQ holds at each  $x \in \Sigma(p)$ , then  $\varphi$  is directionally differentiable at  $p$  and*

$$(4.5) \quad \varphi'(p; h) = \min_{x \in \Sigma(p)} \max_{(y, z) \in \mathcal{M}_p(x)} \nabla_p \mathcal{L}(x, y, z, p)^\top h.$$

Note that if in addition SMCFQ holds, then  $\mathcal{M}_p(x)$  reduces to a singleton  $(y^*, z^*)$  and the directional derivative simplifies to

$$(4.6) \quad \varphi'(p; h) = \min_{x \in \Sigma(p)} \nabla_p \mathcal{L}(x, y^*, z^*, p)^\top h.$$

If instead we suppose GSSOSC holds, the local solution  $x(p)$  is a singleton and the computation of the directional derivative also simplifies.

**THEOREM 4.5** (Theorem 7.5 [23]). *Suppose GSSOSC and MFCQ hold at a stationary solution  $x^*$ . Then near  $p$  the function  $\varphi(\cdot)$  has a finite one-sided directional derivative, and for every  $h \in \mathbb{R}^\ell$ ,*

$$(4.7) \quad \varphi'(p; h) = \max_{(y, z) \in \mathcal{M}_p(x^*)} \nabla_p \mathcal{L}(x^*, y, z, p)^\top h.$$

**4.2. The differentiable case.** Now, we impose more stringent constraint qualifications to recover the setting of the Implicit Function Theorem 3.1: as a consequence the value function  $\varphi$  becomes  $C^2$  differentiable. For a given  $p \in \mathbb{R}^\ell$ , the following results hold not only for global solution  $x \in \Sigma(p)$ , but also *locally* for every KKT stationary point  $w(p) = (x(p), y(p), z(p))$  of NLP( $p$ ). As such, we define the *local value function* as

$$(4.8) \quad \varphi_\ell(p) = f(x(p), p).$$

THEOREM 4.6 (Theorem 3.4.1 [21]). *Suppose SOSC, SCS and LICQ hold at a stationary KKT solution  $(x^*, y^*, z^*)$ . Then the local value function  $\varphi_\ell$  is  $C^2$  near  $p$ , and*

- $\varphi_\ell(p) = \mathcal{L}(x^*, y^*, z^*, p)$ .
- $\nabla_p \varphi_\ell(p) = \nabla_p \mathcal{L}(x^*, y^*, z^*, p)$ .
- and also

$$(4.9) \quad \begin{aligned} \nabla_{pp}^2 \varphi_\ell(p) &= \nabla_p (\nabla_p \mathcal{L}(w^*, p)) = \nabla_{pp}^2 \mathcal{L} + \nabla_{wp}^2 \mathcal{L} J_p w(p), \\ &= \nabla_{pp}^2 \mathcal{L} - \nabla_{wp}^2 \mathcal{L} (\nabla_{ww}^2 \mathcal{L})^{-1} \nabla_{pw}^2 \mathcal{L}. \end{aligned}$$

The previous theorem gives expressions both for the gradient and the Hessian of the optimal value function. It applies to generic nonlinear programs of the form  $\text{NLP}(p)$ . It yields interesting extensions when applied to problem with certain structures.

*Case 1: the constraints are independent of  $p$ .* First, we can prove that if the constraints of the nonlinear problem are independent of the parameter  $p$ ,

$$(4.10) \quad \text{NLP}_{obj}(p) := \min_{x \in \mathbb{R}^n} f(x, p) \quad \text{subject to} \quad \begin{cases} g(x) = 0, \\ h(x) \leq 0, \end{cases}$$

then the gradient and the Hessian depend only on the derivatives of the objective function  $f$ .

COROLLARY 4.7. *Suppose the conditions of Theorem 4.6 hold at a stationary solution  $(x^*, y^*, z^*)$  of problem (4.10). Then, in a neighborhood of  $p$  the local value function  $\varphi_\ell$  is  $C^2$ , with*

- $\varphi_\ell(p) = f(x^*, p)$ .
- $\nabla_p \varphi_\ell(p) = \nabla_p f(x^*)$ .
- $\nabla_{pp}^2 \varphi_\ell(p) = \nabla_{pp}^2 f + (\nabla_{xp}^2 f) J_p x(p)$ .

*Case 2: RHS perturbations and shadow prices.* The other corollary draws a link between Theorem 4.6 and the shadow price of sensitivity: if the parameters appear only in the right-hand-side of the constraint, that is, for  $p \in \mathbb{R}^{m_e}$  and  $q \in \mathbb{R}^{m_i}$ ,

$$(4.11) \quad \text{NLP}_{rhs}(p, q) := \min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \quad \begin{cases} g(x) = p, \\ h(x) \leq q, \end{cases}$$

then the gradient of the value function is equal to the Lagrange multipliers found at the solution. This is well-known result in convex analysis, the problem (4.11) being subject to the so-called *canonical perturbations*.

COROLLARY 4.8 (Optimal value function derivatives for RHS perturbations). *Suppose the conditions of Theorem 4.6 hold at a stationary solution  $(x^*, y^*, z^*)$  of problem (4.11). Then, in a neighborhood of  $(p, q)$  the local value function  $\varphi_\ell$  is  $C^2$ , with*

$$(4.12) \quad \nabla \varphi_\ell(p, q) = \begin{bmatrix} y^* \\ z^* \end{bmatrix} \quad \text{and} \quad \nabla^2 \varphi_\ell(p^*) = \begin{bmatrix} J_p y(p, q) & J_q y(p, q) \\ J_p z(p, q) & J_q z(p, q) \end{bmatrix}.$$

**5. Numerical algorithms for sensitivity analysis.** Now the theory has been laid out, we give an overview of the existing numerical methods for the sensitivity analysis of a nonlinear optimization program (2.1). First, we put a special emphasis on the differentiation of conic programs in §5.1, before focusing on the differentiation of nonlinear programs in §5.2.



**5.1. Differentiation of conic programs.** Conic programming has become an important subfield of optimization. In machine learning, a significant number of problems are conic, motivating a renewed interest for the sensitivity analysis of conic-structured problems. A conic program is a specialization of (2.1), with the following structure

$$(5.1) \quad \min_{x \in \mathbb{R}^n} c^\top x \quad \text{s.t.} \quad Ax = b, \quad x \in \mathcal{K},$$

where  $\mathcal{K}$  is a pointed convex cone. If the cone is the non-negative orthant  $\mathcal{K} = \mathbb{R}_+^n$ , the problem (5.1) becomes a LP. We suppose here that the problem's data are the parameters:  $p = (c, A, b)$ .

The KKT conditions of (5.1) write out

$$(5.2) \quad A^\top y + s = c, \quad Ax = b, \quad x \in K, \quad s \in K^*, \quad x^\top s = 0.$$

As the problem (5.1) is convex, the KKT conditions (5.2) are necessary and sufficient.

*Homogeneous self-dual embedding.* On the contrary to what we exposed in Theorem 3.2, the sensitivity analysis of (5.1) does not differentiate through the KKT conditions (5.2): instead, it uses an equivalent form of (5.2) called the *homogeneous self-dual embedding* (HSD) [9, 11]. The HSD writes out as the problem

$$(5.3) \quad \text{Find } (x, y, \tau, s, \kappa) \quad \text{s.t.} \quad \begin{cases} \begin{bmatrix} 0 & -A^\top & c \\ A & 0 & -b \\ -c^\top & b^\top & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ \tau \end{bmatrix} = \begin{bmatrix} s \\ 0 \\ \kappa \end{bmatrix}, \\ x \in K, \quad s \in K^* \\ (\tau, \kappa) \geq 0 \quad \text{and} \quad x^\top s + \tau \kappa = 0. \end{cases}$$

If  $(x, y, \tau, s, \kappa)$  is solution of (5.3) with  $\tau > 0$  and  $\kappa = 0$ , then  $(x/\tau, y/\tau, s/\tau)$  satisfies the KKT conditions (5.2) and is solution of the conic program (5.1). In general, the parameters  $\tau$  and  $\kappa$  characterize the primal and dual infeasibility of the conic problem (5.1).

We define the skew-symmetric matrix

$$(5.4) \quad Q = \begin{bmatrix} 0 & -A^\top & c \\ A^\top & 0 & -b \\ -c^\top & b^\top & 0 \end{bmatrix},$$

Defining the new cone  $\mathcal{C} = K \times \mathbb{R}^m \times \mathbb{R}_+$  and its dual  $\mathcal{C}^* = K^* \times \{0\}^n \times \mathbb{R}_+$ , the problem (5.3) becomes equivalent to

$$(5.5) \quad \text{Find } (u, v) \in \mathcal{C} \times \mathcal{C}^* \quad \text{s.t.} \quad Qu = v, \quad u^\top v = 0.$$

*Residual map.* Let  $z \in \mathbb{R}^{n+m+1}$ . Using Moreau's decomposition theorem, we know that  $z = u - v$  with  $(u, v) \in \mathcal{C} \times \mathcal{C}^*$  and  $u^\top v = 0$  if and only if  $u = P_{\mathcal{C}}(z)$  and  $v = -P_{\mathcal{C}^o}(z)$ .

As a consequence, the HSD (5.5) rewrites more compactly as finding  $z \in \mathbb{R}^{n+m+1}$  such that

$$(5.6) \quad -P_{\mathcal{C}^o}(z) = QP_{\mathcal{C}}(z).$$

The residual map  $R(z)$  returns the HSD residual:

$$(5.7) \quad R(z) = QP_{\mathcal{C}}(z) + P_{\mathcal{C}^o}(z) = ((Q - I)P_{\mathcal{C}} + I)z.$$

If we decompose  $z^*$  component by component as  $z^* = (u^*, v^*, w^*)$ , with  $(u^*, v^*, w^*) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}_+^*$ , the KKT solutions  $(x^*, y^*, s^*)$  of (5.1) is recovered as

$$(5.8) \quad (x^*, y^*, s^*) = \phi(z^*) \quad \text{with} \quad \phi(z^*) := \frac{1}{w^*} (P_{\mathcal{C}}(z^*), -P_{\mathcal{C}^\circ}(z^*), w^*) .$$

*Sensitivity analysis.* We now have all the elements to evaluate the sensitivity at a given KKT primal-dual solution  $w^* = (x^*, s^*, y^*)$  of (5.2). By defining  $u^* = (x^*, y^*, 1)$ ,  $v^* = (s^*, 0, 0)$  and  $z^* = u^* - v^*$ , the vector  $z^*$  is solution of the residual map:  $R(z^*) = 0$ . If the projection operator  $P_{\mathcal{C}}$  is differentiable, then  $R$  is itself differentiable with  $J_z R(z) = (Q + I)J_z P_{\mathcal{C}}(z) - I$ . In that case, by applying the implicit function theorem 3.1 to  $R(z) = 0$ , we get the Jacobian

$$(5.9) \quad J_p z^* = -(J_z R(z^*))^{-1} J_p R(z^*) .$$

We recover the sensitivity of the primal-dual solution by noting that  $w^* = \phi(z^*)$ : after applying the chain-rule, we get  $J_p w^* = J_z \phi \cdot J_p z^*$ .

Note that in general the projection operator  $P_K$  is not differentiable. However, as recalled in [2], if  $K$  encodes the nonnegative orthant, the second-order cone or the positive definite cone then  $P_K$  is strongly semismooth (see Appendix A.3), hence differentiable almost everywhere. As a result, the residual mapping (5.7) is a strongly semismooth function: the equation (5.9) is well-defined if we use a version of the implicit function theorem adapted for the strongly semismooth case (and involving Clarke generalized Jacobians). Another rigorous treatment has been made in [9], based on the newly introduced concept of *conservative Jacobians*.

*Implementation.* With the emergence of the differentiable programming paradigm, numerous packages have been developed to differentiate through conic program (5.1). The approach has been implemented in `cvxpylayer` [1] and `DiffOpt.jl` [51]. We emphasize that factorizing the matrix  $\nabla_z R$  in (5.9) can be challenging, as on the contrary to the KKT matrix the Jacobian  $\nabla_z R$  is non symmetric. For that reason, the system (5.9) is solved either using a LU factorization, or using a Krylov method (the LSQR algorithm accommodates nicely the degenerate case where the Jacobian of the residual  $\nabla_z R$  is non-invertible).

**5.2. Differentiation of nonlinear programs.** The solution of  $\text{NLP}(p)$  has been studied by Fiacco and McCormick in their seminal work [25]: the main idea is to reformulate the problem with penalties and solve the resulting unconstrained problem using Newton method. In that case, the sensitivity analysis uses the approximate solution returned by the algorithm to compute the directional derivative  $x'(p, h)$ .

**5.2.1. Fiacco and SENSUMT.** We recall the penalty approach described originally in [22]. Although this approach is now slightly outdated, it introduces some fundamental ideas still being used today. We follow the approach described in [21, Chapter 6]. The idea is to reformulate  $\text{NLP}(p)$  with inner and outer penalties. For a strictly feasible point  $x$  (in the sense  $h(x, p) < 0$ ), Fiacco and McCormick define the functional

$$(5.10) \quad W(x, r, p) = f(x, p) - r \sum_{i=1}^{m_i} \log(-h_i(x, p)) + \frac{1}{2r} \sum_{j=1}^{m_e} g_j^2(x, p) ,$$

where  $r > 0$  is a penalty parameter appearing in the barrier associated to the inequality constraints and the quadratic penalty associated to the equality equations.

For a fixed penalty  $r$ , a stationary point  $\nabla_x W(x(p, r), r, p) = 0$  is found using the traditional Newton method. SENSUMT [22] was developed as an extension of SUMT to compute the sensitivity using the penalty function  $W(\cdot)$ .

Note that the gradient of the penalty  $W(\cdot)$  and the gradient of the Lagrangian (2.5) are defined respectively as

$$(5.11) \quad \begin{aligned} \nabla_x W(x, r, p) &= \nabla_x f(x, p) + \sum_{i=1}^{m_i} (r/h_i) \nabla_x h_i(x, p) + \sum_{j=1}^{m_e} (g_j/r) \nabla_x g_j(x, p) , \\ \nabla_x \mathcal{L}(x, y, z, p) &= \nabla_x f(x, p) + \sum_{i=1}^{m_i} z_i \nabla_x h_i(x, p) + \sum_{j=1}^{m_e} y_j \nabla_x g_j(x, p) , \end{aligned}$$

where we defined  $h_i := h_i(x, p)$  and  $g_j := g_j(x, p)$ . We observe that for any solution  $x^*$  of (5.10), the multipliers  $z^* = r/h(x^*, p)$  and  $y^* = g(x^*, p)/r$  satisfy the stationary condition  $\nabla_x \mathcal{L}(x^*, y^*, z^*, p) = 0$ : The variables  $(y^*, z^*)$  are approximations of the optimal Lagrange multipliers, both depending on the penalty  $r$ . The penalty  $r$  is here interpreted as a new parameter, and is valid candidate to apply Theorem 3.2.

**PROPOSITION 5.1** (Theorem 6.2.1 [21]). *Let  $p^* \in \mathbb{R}^\ell$ . Suppose that a stationary solution  $x^*$  of NLP( $p$ ) LICQ, SOSC and SCS hold. Then in a neighborhood of  $(0, p^*)$  there exists a unique differentiable function  $w(r, p) = (x(r, p), y(r, p), z(r, p))$  satisfying*

$$(5.12) \quad \begin{aligned} \nabla_x \mathcal{L}(x, y, z, p) &= 0 , \\ g_j(x, p) &= y_j r \quad \forall j \in [m_e] , \\ z_i h_i(x, p) &= r , \quad \forall i \in [m_i] , \end{aligned}$$

and such that  $w(0, p^*) = (x^*, y^*, z^*)$ . In addition, for any  $(r, p)$  near  $(0, p^*)$ ,  $x(r, p)$  is locally unique unconstrained local minimizing point of  $W(x, r, p)$  with  $h_i(x(r, p), p) < 0$  and  $\nabla_{xx}^2 W(x(r, p), r, p)$  positive definite.

At the solution of (5.10), the sensitivity of the primal  $x(r, p)$  is given by the Implicit Function Theorem 3.1 as

$$(5.13) \quad J_p x(r, p) = -(\nabla_{xx}^2 W)^{-1} \nabla_{pw} W ,$$

involving only the inverse of the  $n \times n$  Hessian matrix  $\nabla_{xx}^2 W$ . Using Proposition 5.1, we get that  $\lim_{r \rightarrow 0} J_p x(r, p) = J_p x^*(p)$ . The dual sensitivities are recovered with (5.12) as  $z_i(r, p) = r/h_i(x(r, p), p)$  and  $y_j = g_j(x(r, p), p)/r$ , giving the respective Jacobians

$$(5.14) \quad \begin{aligned} J_p y(r, p) &= \frac{1}{r} (J_x g(x(r, p), p) \cdot J_p x(r, p) + J_p g(x(r, p), p)) , \\ J_p z(r, p) &= -\frac{r}{h(x(r, p), p)} (J_x h(x(r, p), p) \cdot J_p x(r, p) + J_p h(x(r, p), p)) . \end{aligned}$$

**5.2.2. Sensitivity in interior-point.** The SUMT algorithm described in the previous paragraph has fallen out of fashion, as the resulting problem (5.10) becomes highly ill-conditioned when we are approaching an optimal solution. However, SUMT has nurtured the development of interior-point methods in the 1980s: the primal-dual interior-point method (IPM) is now one of the most widely used algorithm to solve nonlinear program NLP( $p$ ).

The solution sensitivity of primal-dual interior point follows from the same principle as in Theorem 5.1. As an example, we describe the approach uses in sIpopt [42], a package for the sensitivity analysis of NLP developed as an extension of the primal-dual interior-point solver Ipopt. Ipopt rewrites the inequalities in  $\text{NLP}(p)$  as equality constraints by introducing additional slack variables, giving a reformulated NLP problem with only bound constraints  $x \geq 0$ :

$$(5.15) \quad \begin{aligned} & \min_x f(x, p) \\ \text{s.t.} \quad & g(x, p) = 0, \quad x \geq 0. \end{aligned}$$

For a given barrier parameter  $\mu > 0$ , the bound constraints  $x \geq 0$  are penalized using a barrier term in the objective function:

$$(5.16) \quad \begin{aligned} & \min_x f(x, p) - \mu \sum_{i=1}^n \log(x_i) \\ \text{s.t.} \quad & g(x, p) = 0. \end{aligned}$$

Ipopt solves a sequence of subproblems (5.16) for given barrier parameters  $\{\mu_k\}_k$ , with  $\mu_k \rightarrow 0$ . For a fixed  $\mu$ , solving (5.16) with primal-dual IPM amounts to solve the system of nonlinear equations  $F_\mu(x, y, z, p) = 0$ , with

$$(5.17) \quad F_\mu(x, y, z, p) = \begin{bmatrix} \nabla_x f(x, p) + J_x g(x, p)^\top y - z \\ g(x, p) \\ Xz - \mu e \end{bmatrix},$$

where we have defined  $X = \text{diag}(x)$ .

The following proposition is a corollary of Proposition 5.1.

**PROPOSITION 5.2** (Convergence of interior-point). *Let  $w^* = (x^*, y^*, z^*)$  be a local primal-dual stationary point of (5.15). Suppose that LICQ, SCS and SOSC hold at  $w^*$ . We denote by  $x(\mu_k, p)$  the solution of the barrier problem (5.16) at iteration  $k$ . Then*

- *There is at least one subsequence of  $x(\mu_k, p)$  converging to  $x^*$  and for every convergent subsequence, the corresponding barrier multiplier approximation is bounded and converges to  $(y^*, z^*)$ .*
- *There exists a unique continuously differentiable function  $w(\cdot, p)$  existing in a neighborhood of 0 for  $\mu > 0$  and solution of (5.16), and such that*

$$(5.18) \quad \lim_{\mu \rightarrow 0} w(\mu, p) = w^*.$$

- *For  $\mu$  small enough,*

$$(5.19) \quad J_p w(\mu, p) = -(J_w F_\mu)^{-1} J_p F_\mu.$$

To solve the sparse linear system (5.19), sIpopt uses the factorization of the KKT system computed by Ipopt at the local solution  $w(\mu, p)$ . To handle change in the active set, sIpopt uses a fix-and-relax strategy, without resorting to the resolution of the QP (3.17). The new constraints activated are incorporated in the linear system (5.19), and the system is solved using a Schur-complement approach.

*Variant 1: CasADi.* CasADi [4] is a powerful modeler used to solve nonlinear programs (2.1). The sensitivity analysis implemented in CasADi is slightly different than sIpopt [5]. Once a solution returned by the optimization solver, CasADi applies a primal-dual active set method to determine exactly the active set at the solution (an information that interior-point does not return by default, unless a crossover is used). Then, CasADi uses a custom sparse QR solver to solve a non-symmetric formulation of the KKT condition. Such QR factorization allows to quickly detect and resolve the singularities associated to potential degeneracy.

*Variant 2: QPTH.* sIpopt's method has been recently specialized in [3] to the parallel extraction of the sensitivity for a quadratic program (QP), for a fixed number of parameters  $(p_1, \dots, p_N)$ . The solver QPTH solves  $N$  QP problems in parallel using a specialized interior-point method leveraging batched dense linear algebra on the GPU. Once a solution found, the sensitivity are evaluated using a formula analogous to (5.19). This has been proven relevant for machine learning applications, where neural network are trained with batched stochastic gradient descent. To the best of our knowledge, the method has not been extended to sparse problems, limiting its adoption beyond machine learning. Also, little emphasis is put on the degenerate case.

**5.3. Path following method.** In nonlinear programming, one of the main application for sensitivity analysis is updating the solution of  $\text{NLP}(p)$  after a small change in the parameter  $p$ : once the directional derivative is obtained, we can compute an approximation of a new primal solution  $x(p')$  by setting the direction  $h = p' - p$  and using the Taylor expansion [58]:

$$(5.20) \quad x(p') = x(p) + x'(p, h) + o(\|h\|) .$$

However, the method is valid only for small perturbations around  $p^*$ , without any active set change. In case active set changes are occurring, they have to be accommodated locally using a path following method,

A path following method tracks the path of optimal solution between  $p$  and the new parameter  $p'$ . To do so, it defines the affine interpolation  $p(t) = p + t(p' - p)$  for  $t \in [0, 1]$  and tracks the solution of  $\text{NLP}(p)$  between  $t = 0$  and  $t = 1$ . This approach returns more accurate approximations, notably when facing active set changes. For a sequence of scalars  $t_{(0)} < t_{(1)} < \dots < t_{(N)}$  such that  $t_{(0)} = 0$  and  $t_{(N)} = 1$ , the perturbation between  $t_{(m)}$  and  $t_{(m+1)}$  is defined as

$$(5.21) \quad h_{(m)} = (t_{(m+1)} - t_{(m)})(p' - p) , \quad \forall m = 0, \dots, N-1 .$$

Then, we apply  $N - 1$  sensitivity updates with the direction  $h_{(m)}$  to follow the path of the primal-dual solution  $w(t)$  between  $t_{(0)}$  and  $t_{(N)}$ .

We refer to [58] for a description of the path following method. The method has been extended later on [34] for degenerate problems where the dual solution is non-unique, using the method of Ralph and Dempe described in §3.4. The algorithm proposed in [34] tracks the active set changes by updating iteratively the dual multipliers using the LP (3.18). As such, it can be interpreted as running  $N$  iterations of a sequential quadratic programming (SQP) algorithm. The original method [34] didn't come with any convergence guarantee, but it has been extended later in [37] with a rigorous convergence proof. In [37], the path following method is interpreted as a predictor-corrector method, where the corrector solves a linear system to update the sensitivity and the predictor solves a QP to track the active set changes. The

multipliers are again updated by solving a LP equivalent to (3.18). We note that the method requires the evaluation of the second-order derivatives of each constraint in order to solve the QP in the predictor step. To the best of our knowledge, we are not aware of a generic package that implements the method described in [37].

**Acknowledgements.** We thank Léonard Moracchini for his help in establishing the first version of this manuscript, and Andrew Rosemberg and Tristan Rigaut for their helpful feedbacks.

#### REFERENCES

- [1] A. AGRAWAL, B. AMOS, S. BARRATT, S. BOYD, S. DIAMOND, AND J. Z. KOLTER, *Differentiable convex optimization layers*, Advances in neural information processing systems, 32 (2019).
- [2] A. ALI, E. WONG, AND J. Z. KOLTER, *A semismooth newton method for fast, generic convex programming*, in International Conference on Machine Learning, PMLR, 2017, pp. 70–79.
- [3] B. AMOS AND J. Z. KOLTER, *Optnet: Differentiable optimization as a layer in neural networks*, in International Conference on Machine Learning, PMLR, 2017, pp. 136–145.
- [4] J. A. ANDERSSON, J. GILLIS, G. HORN, J. B. RAWLINGS, AND M. DIEHL, *CasADi: a software framework for nonlinear optimization and optimal control*, Mathematical Programming Computation, 11 (2019), pp. 1–36.
- [5] J. A. ANDERSSON AND J. B. RAWLINGS, *Sensitivity analysis for nonlinear programming in CasADi*, IFAC-PapersOnLine, 51 (2018), pp. 331–336.
- [6] P. I. BARTON, K. A. KHAN, P. STECHLINSKI, AND H. A. WATSON, *Computationally relevant generalized derivatives: theory, evaluation and applications*, Optimization Methods and Software, 33 (2018), pp. 1030–1072.
- [7] M. BENZI, G. H. GOLUB, AND J. LIESEN, *Numerical solution of saddle point problems*, Acta numerica, 14 (2005), pp. 1–137.
- [8] M. BLONDEL, Q. BERTHET, M. CUTURI, R. FROSTIG, S. HOYER, F. LLINARES-LÓPEZ, F. PEDREGOSA, AND J.-P. VERT, *Efficient and modular implicit differentiation*, Advances in neural information processing systems, 35 (2022), pp. 5230–5242.
- [9] J. BOLTE, T. LE, E. PAUWELS, AND T. SILVETI-FALLS, *Nonsmooth implicit differentiation for machine-learning and optimization*, Advances in neural information processing systems, 34 (2021), pp. 13537–13549.
- [10] J. F. BONNANS AND A. SHAPIRO, *Perturbation analysis of optimization problems*, Springer Science & Business Media, 2000.
- [11] E. BUSSETI, W. M. MOURSİ, AND S. BOYD, *Solution refinement at regular points of conic problems*, Computational Optimization and Applications, 74 (2019), pp. 627–643.
- [12] J. M. DANSKIN, *The theory of max-min and its application to weapons allocation problems*, vol. 5, Springer Science & Business Media, 1967.
- [13] V. DEMIGUEL AND F. J. NOGALES, *On decomposition methods for a class of partially separable nonlinear programs*, Mathematics of Operations Research, 33 (2008), pp. 119–139.
- [14] S. DEMKO, W. F. MOSS, AND P. W. SMITH, *Decay rates for inverses of band matrices*, Mathematics of computation, 43 (1984), pp. 491–499.
- [15] S. DEMPE, *Directional differentiability of optimal solutions under Slater’s condition*, Mathematical programming, 59 (1993), pp. 49–69.
- [16] ———, *Foundations of bilevel programming*, Springer Science & Business Media, 2002.
- [17] A. L. DONTCHEV AND R. T. ROCKAFELLAR, *Implicit functions and solution mappings: A view from variational analysis*, vol. 11, Springer, 2009.
- [18] F. FACCHINEI AND J.-S. PANG, *Finite-dimensional variational inequalities and complementarity problems*, Springer, 2003.
- [19] A. V. FIACCO, *Sensitivity analysis for nonlinear programming using penalty methods*, Mathematical programming, 10 (1976), pp. 287–311.
- [20] ———, *Optimal value continuity and differential stability bounds under the mangasarian-fromovitz constraint qualification*, Mathematical programming with data perturbations II, (1982), pp. 65–90.
- [21] ———, *Introduction to sensitivity and stability analysis in non linear programming*, New York: Academic Press, 1983.
- [22] A. V. FIACCO AND A. GHAEMI, *A user’s manual for SENSUMT: A penalty function computer program for solution, sensitivity analysis, and optimal value bound calculation in parametric nonlinear programs.*, tech. rep., 1980.

- [23] A. V. FIACCO AND J. KYPARISIS, *Sensitivity analysis in nonlinear programming under second order assumptions*, in Systems and Optimization: Proceedings of the Twente Workshop Enschede, The Netherlands, April 16–18, 1984, Springer, 1984, pp. 74–97.
- [24] A. V. FIACCO AND J. LIU, *Degeneracy in nlp and the development of results motivated by its presence*, Annals of Operations Research, 46 (1993), pp. 61–80.
- [25] A. V. FIACCO AND G. P. MCCORMICK, *Nonlinear programming: sequential unconstrained minimization techniques*, SIAM, 1968.
- [26] J. GAUVIN, *A necessary and sufficient regularity condition to have bounded multipliers in non-convex programming*, Mathematical Programming, 12 (1977), pp. 136–138.
- [27] ———, *The method of parametric decomposition in mathematical programming: the nonconvex case*, in IIASA PROCEEDINGS SERIES, 1978, p. 131.
- [28] J. GAUVIN AND F. DUBEAU, *Differential properties of the marginal function in mathematical programming*, Optimality and Stability in Mathematical Programming, (1982), pp. 101–119.
- [29] J. C. GILBERT, *Automatic differentiation and iterative processes*, Optimization methods and software, 1 (1992), pp. 13–21.
- [30] A. GRIEWANK AND C. FAURE, *Reduced functions, gradients and hessians from fixed-point iterations for state equations*, Numerical Algorithms, 30 (2002), pp. 113–139.
- [31] A. GRIEWANK AND A. WALTHER, *Evaluating derivatives: principles and techniques of algorithmic differentiation*, SIAM, 2008.
- [32] W. W. HOGAN, *Point-to-set maps in mathematical programming*, SIAM review, 15 (1973), pp. 591–603.
- [33] A. F. IZMAILOV AND M. V. SOLODOV, *A note on solution sensitivity for karush–kuhn–tucker systems*, Mathematical Methods of Operations Research, 61 (2005), pp. 347–363.
- [34] J. JÄSCHKE, X. YANG, AND L. T. BIEGLER, *Fast economic model predictive control based on NLP-sensitivities*, Journal of Process Control, 24 (2014), pp. 1260–1272.
- [35] K. JITTORNTUM, *Solution point differentiability without strict complementarity in nonlinear programming*, Sensitivity, Stability and Parametric Analysis, (1984), pp. 127–138.
- [36] M. KOJIMA, *Strongly stable stationary solutions in nonlinear programs*, in Analysis and computation of fixed points, Elsevier, 1980, pp. 93–138.
- [37] V. KUNGURTSOV AND J. JÄSCHKE, *A predictor-corrector path-following algorithm for dual-degenerate parametric optimization problems*, SIAM Journal on Optimization, 27 (2017), pp. 538–564.
- [38] J. KYPARISIS, *On uniqueness of Kuhn-Tucker multipliers in nonlinear programming*, Mathematical Programming, 32 (1985), pp. 242–246.
- [39] ———, *Sensitivity analysis for nonlinear programs and variational inequalities with nonunique multipliers*, Mathematics of operations research, 15 (1990), pp. 286–298.
- [40] R. LÓPEZ-NEGRETÉ AND L. T. BIEGLER, *A moving horizon estimator for processes with multi-rate measurements: A nonlinear programming sensitivity approach*, Journal of Process Control, 22 (2012), pp. 677–688.
- [41] Y. NESTEROV, *Lexicographic differentiation of nonsmooth functions*, Mathematical programming, 104 (2005), pp. 669–700.
- [42] H. PIRNAY, R. LÓPEZ-NEGRETÉ, AND L. T. BIEGLER, *Optimal sensitivity based on IPOPT*, Mathematical Programming Computation, 4 (2012), pp. 307–331.
- [43] D. RALPH AND S. DEMPE, *Directional derivatives of the solution of a parametric nonlinear program*, Mathematical programming, 70 (1995), pp. 159–172.
- [44] S. M. ROBINSON, *Generalized equations and their solutions, Part I: Basic theory*, Springer, 1979.
- [45] ———, *Strongly regular generalized equations*, Mathematics of Operations Research, 5 (1980), pp. 43–62.
- [46] ———, *Generalized equations and their solutions, part II: applications to nonlinear programming*, Springer, 1982.
- [47] S. SCHOLTES, *Introduction to piecewise differentiable equations*, Springer Science & Business Media, 2012.
- [48] A. SHAPIRO, *Second order sensitivity analysis and asymptotic theory of parametrized nonlinear programs*, Mathematical Programming, 33 (1985), pp. 280–299.
- [49] ———, *On differential stability in stochastic programming*, Mathematical Programming, 47 (1990), pp. 107–116.
- [50] ———, *Asymptotic analysis of stochastic programs*, Annals of Operations Research, 30 (1991), pp. 169–186.
- [51] A. SHARMA, M. BESANÇON, J. D. GARCIA, AND B. LEGAT, *Flexible differentiable optimization via model transformations*, arXiv preprint arXiv:2206.06135, (2022).

- [52] S. SHIN, M. ANITESCU, AND V. M. ZAVALA, *Overlapping schwarz decomposition for constrained quadratic programs*, in 2020 59th IEEE Conference on Decision and Control (CDC), IEEE, 2020, pp. 3004–3009.
- [53] ———, *Exponential decay of sensitivity in graph-structured nonlinear programs*, SIAM Journal on Optimization, 32 (2022), pp. 1156–1183.
- [54] P. STECHLINSKI, J. JÄSCHKE, AND P. I. BARTON, *Generalized sensitivity analysis of nonlinear programs using a sequence of quadratic programs*, Optimization, 68 (2019), pp. 485–508.
- [55] P. STECHLINSKI, K. A. KHAN, AND P. I. BARTON, *Generalized sensitivity analysis of nonlinear programs*, SIAM Journal on Optimization, 28 (2018), pp. 272–301.
- [56] L. F. VALENZUELA, R. BROWN, AND M. PAVONE, *Decentralized implicit differentiation*, arXiv preprint arXiv:2403.01260, (2024).
- [57] V. M. ZAVALA AND M. ANITESCU, *Real-time nonlinear optimization as a generalized equation*, SIAM Journal on Control and Optimization, 48 (2010), pp. 5444–5467.
- [58] V. M. ZAVALA AND L. T. BIEGLER, *The advanced-step nmqc controller: Optimality, stability and robustness*, Automatica, 45 (2009), pp. 86–93.

**Appendix A. Derivatives.** Let  $f : \Omega \subset E \rightarrow F$  a function, with  $E$  and  $F$  two normed spaces. We start to recall the usual derivatives, before introducing the Clarke Jacobian more adapted in the nonsmooth context.

### A.1. Differentiability.

DEFINITION A.1 (Directional derivatives). *Let  $x \in \Omega$ . The function  $f$  has a directional derivative at  $x$  in the direction  $h \in E$  if  $x + th \in \Omega$  for  $t$  small enough, and if the following limit exists*

$$(A.1) \quad f'(x; h) = \lim_{t \downarrow 0} \frac{f(x + th) - f(x)}{t} .$$

DEFINITION A.2 (Gâteaux-differentiability). *The function  $f$  is Gâteaux differentiable (or  $G$ -differentiable) in  $x \in \Omega$  if there exists a directional derivatives for every direction  $h \in E$ , and if the application*

$$(A.2) \quad h \rightarrow f'(x; h) ,$$

*is continuous and linear. We note  $f'(x)$  the operator such that  $f'(x; h) = f'(x) \cdot h$ .*

We note that a  $G$ -differentiable function is not necessarily continuous.

DEFINITION A.3 (Fréchet-differentiability). *The function  $f$  is Fréchet differentiable ( $F$ -differentiable, or simply differentiable) at  $x \in \Omega$  if there exists a bounded linear operator  $L$  such that*

$$(A.3) \quad \lim_{\|h\| \downarrow 0} \frac{1}{\|h\|} \left( f(x + h) - f(x) - Lh \right) = 0 .$$

$F$ -differentiability implies that  $f$  is continuous. The operator  $L$  is the *derivative* of  $f$  in  $x$ . The condition (A.3) rewrites equivalently

$$(A.4) \quad f(x + h) = f(x) + Lh + o(\|h\|) .$$

The Hadamard derivative gives the directional derivative along a curve tangential to  $h$ . It is required to study the sensitivity of infinite-dimensional problems, as often encountered in statistics or stochastic optimization.

DEFINITION A.4 (Hadamard-differentiability). *The function  $f$  is Hadamard differentiable at  $x$  if for any mapping  $\varphi : \mathbb{R}_+ \rightarrow X$  such that  $\varphi(0) = x$  and  $\frac{1}{t}(\varphi(t) - \varphi(0))$  converges to a vector  $h$  as  $t \downarrow 0$ , the limit*

$$(A.5) \quad d_H f(x; h) = \lim_{t \downarrow 0} \frac{1}{t} (f(\varphi(t)) - f(x)) ,$$



does exist.

For any sequences  $\{h_n\}_n$  and  $\{t_n\}_n$  such that  $h_n \rightarrow h$  and  $t_n \rightarrow 0^+$ , the Hadamard directional derivative can also be written in the form

$$(A.6) \quad d_H f(x; h) = \lim_{n \rightarrow \infty} \frac{1}{t_n} (f(x + t_n h_n) - f(x)) .$$

The Hadamard derivative has the advantage of being compatible with the chain-rule with minimal assumptions.

**PROPOSITION A.5 (Chain-rule).** *Let  $f$  be Hadamard directionally differentiable at  $x$ , and  $g$  be Hadamard directionally differentiable at  $f(x)$ . Then the composite mapping  $g \circ f$  is Hadamard directionally differentiable at  $x$ , and satisfies the chain-rule  $d_H(g \circ f)(x; h) = d_H g(f(x); d_H f(x; y))$ .*

Whereas the Hadamard derivative allows for different rates for every directions, the Fréchet derivative imposes the rate is the same for each direction  $h$ . For finite-dimensional space  $E$ , if the directional derivative is continuous then the Fréchet and the Hadamard derivatives coincide.

**A.2. Generalized differentiability.** The notion of derivatives can be extended to nonsmooth functions. We recall the following important results.

**THEOREM A.6 (Rademacher Theorem).** *Let  $f : \Omega \rightarrow F$  be a locally Lipschitz continuous function. Then  $f$  is differentiable almost everywhere in the sense of the Lebesgue measure.*

The concept of Dini-derivatives generalizes the notion of directional differentiability for non-smooth functions.

**DEFINITION A.7 (Dini-derivatives).** *Let  $x \in \Omega$ . The upper Dini derivative in the direction  $h$  is defined as*

$$(A.7) \quad D^+ f(x; h) = \limsup_{t \downarrow 0} \left( \frac{f(x + th) - f(x)}{t} \right) .$$

Accordingly, the lower Dini derivative is defined as

$$(A.8) \quad D^- f(x; h) = \liminf_{t \downarrow 0} \left( \frac{f(x + th) - f(x)}{t} \right) .$$

**DEFINITION A.8 (Bouligand-differential).** *Let  $\mathcal{D}_f$  be the set of points at which  $f$  is (Fréchet) differentiable. The B-differential of  $f$  at  $x \in \Omega$  is the set defined as*

$$(A.9) \quad \partial_B f(x) = \{ J \in \mathcal{L}(E, F) : \exists \{x_k\}_k \in \mathcal{D}_f \text{ such that } x_k \rightarrow x, f'(x_k) \rightarrow J \} .$$

**DEFINITION A.9 (Clarke-differential).** *The C-differential of  $f$  at  $x \in \Omega$  is the convex hull of the Bouligand-differential, that is*

$$(A.10) \quad \partial_C f(x) = \text{conv } \partial_B f(x) .$$

**DEFINITION A.10 (Clarke-generalized directional derivative).** *The Clarke-generalized directional derivative of  $f$  in the direction  $h$  is defined by*

$$(A.11) \quad d_C f(x; h) = \limsup_{y \rightarrow x} \frac{1}{t} [f(y + th) - f(y)] .$$

PROPOSITION A.11. *Suppose that  $f$  is  $L$ -Lipschitz near  $x \in \Omega$ . Then*

1.  $\partial_C f(x)$  is nonempty compact and convex,
2.  $\partial_C f(x)$  is locally bounded and upper semi-continuous at  $x$ .
3.  $d_C f(x; \cdot)$  is the support function of  $\partial_C f(x)$ , that is,

$$(A.12) \quad d_C f(x; h) = \max\{s^\top h : \forall s \in \partial_C f(x)\}.$$

PROPOSITION A.12. *If  $f$  is  $G$ -differentiable at  $x$ , then  $f'(x) \in \partial_C f(x)$ . If in addition  $f$  is  $C^1$  at  $x$ , then  $\partial_C f(x) = \{f'(x)\}$ .*

PROPOSITION A.13 (Chain-rule). *Suppose that  $f$  is  $L$ -Lipschitz continuous on  $\Omega$ . If the function  $g : F \rightarrow G$  is Lipschitz-continuous near  $f(x)$ , then*

$$(A.13) \quad \partial_C(g \circ f)(x) \subset \text{conv}\{GF : G \in \partial_C g(f(x)), F \in \partial_C f(x)\}.$$

### A.3. Semi-smooth functions.

DEFINITION A.14 (Semi-smoothness). *The function  $f : \Omega \rightarrow F$  is semi-smooth in  $x \in \Omega$  if:*

1.  $f$  is  $L$ -Lipschitz near  $x$ ;
2.  $f$  has directional derivatives at  $x$  in all directions;
3. When  $h \rightarrow 0$ ,  $\sup_{J \in \partial_C f} \|f(x+h) - f(x) - Jh\| = o(\|h\|)$ .

PROPOSITION A.15. *Suppose  $f$  is  $L$ -Lipschitz near  $x$ , and admits directional derivatives at  $x$  in all directions. Then the following statements are equivalent:*

1.  $f$  is semi-smooth ;
2. for  $h \rightarrow 0$ ,  $\sup_{J \in \partial_C f} \|Jh - f'(x; h)\| = o(\|h\|)$ .
3. for  $h \rightarrow 0$  such that  $x+h \in \mathcal{D}_f$ ,  $f'(x+h)h - f'(x; h) = o(\|h\|)$ .

**A.4.  $PC^1$  functions.** The class of  $PC^1$  functions is a specific tool in nonsmooth analysis, often employed for the sensitivity analysis of nonlinear programs.

DEFINITION A.16. *The function  $f : X \rightarrow F$  is piecewise-differentiable ( $PC^1$ ) at  $x \in X$  if there exists a neighborhood  $N \subset X$  and a finite collection of  $C^1$  selection functions  $\mathcal{F}_f(x) = \{f_{(1)}, \dots, f_{(k)}\}$  defined on  $N$  such that  $f$  continuous on  $N$  and for all  $y \in N$ ,  $f(y) \in \{f_{(i)}(y)\}_{i \in [k]}$ .*

If  $f$  is  $PC^1$ , we define the active set associated at  $x \in X$  as

$$(A.14) \quad \mathcal{I}_f^{ess}(x) := \{i \in [k] : \forall y \in N, f(y) = f_{(i)}(y)\}.$$

PROPOSITION A.17. *Let  $f : X \rightarrow F$  a  $PC^1$  function at  $x \in X$ . Then*

- $\mathcal{I}_f^{ess}(x)$  is a non-empty set;
- $f$  is Lipschitz continuous near  $x$  and directionally differentiable at  $x$ , with

$$(A.15) \quad \partial_B f(x) = \{\nabla f_{(i)}(x) : i \in \mathcal{I}_f^{ess}(x)\}.$$

*In addition, the directional derivative  $f'(x; \cdot)$  is piecewise linear.*

The notion of *coherent orientation* plays a central role in obtaining the invertibility of a  $PC^1$  function in order to apply the implicit function theorem.

DEFINITION A.18 (Coherent orientation). *Let  $U \subset \mathbb{R}^n$  and  $V \subset \mathbb{R}^m$  be open. A  $PC^1$  function  $F : U \times W \rightarrow \mathbb{R}^n$  is said coherently oriented with respect to  $x$  at  $(x^*, y^*) \in U \times V$  if all matrices in  $\partial_{B,x} f(x^*, y^*)$  have the same non-vanishing determinant sign.*

**A.5. Lexicographic derivatives.** Lexicographic derivatives have been introduced by Nesterov in [41] as the derivatives of *lexicographically smooth* (L-smooth) functions.

DEFINITION A.19 (Lexicographically smooth function). *Let  $f : X \rightarrow \mathbb{R}^m$  a locally Lipschitz continuous function on  $X$ .  $f$  is lexicographically smooth at  $x$  if for any  $k \in \mathbb{N}$  and any  $M = [m_{(1)}, \dots, m_{(k)}] \in \mathbb{R}^{n \times k}$  the following higher-order derivatives are well defined*

$$(A.16) \quad \begin{aligned} f_{x,M}^{(0)} : \mathbb{R}^n &\rightarrow \mathbb{R}^m : d \rightarrow f'(x; d) \\ f_{x,M}^{(j)} : \mathbb{R}^n &\rightarrow \mathbb{R}^m : d \rightarrow [f_{x,M}^{(j-1)}]'(m_{(j)}; d) \quad \forall j = 1, \dots, k \end{aligned}$$

The class of L-smooth functions is closed under composition.  $C^1$  functions, convex functions and  $PC^1$  functions are all L-smooth functions.

DEFINITION A.20 (Lexicographic derivative). *Let  $f : X \rightarrow \mathbb{R}^m$  a L-smooth function at  $x$ , and  $M \in \mathbb{R}^{n \times n}$  a nonsingular matrix. The lexicographic derivative of  $f$  at  $x$  in the direction  $M$  is defined as*

$$(A.17) \quad d_L f(x; M) = \nabla f_{x,M}^{(n)}(0) \in \mathbb{R}^{m \times n}.$$

The lexicographic subdifferential of  $f$  at  $x$  is defined as

$$(A.18) \quad \partial_L f(x) = \{d_L f(x; N) : N \in \mathbb{R}^{n \times n}, \det(N) \neq 0\}.$$

For any  $k \in \mathbb{N}$ ,  $M = [m_{(1)}, \dots, m_{(k)}] \in \mathbb{R}^{n \times k}$  the LD-derivative of  $f$  in the directions  $M$  is defined as

$$(A.19) \quad f'(x; M) = [f_{x,M}^{(0)}(m_{(1)}), f_{x,M}^{(1)}(m_{(2)}), \dots, f_{x,M}^{(k-1)}(m_{(k)})].$$

The LD-derivative is uniquely defined for  $M \in \mathbb{R}^{n \times k}$ . If in addition  $M \in \mathbb{R}^{n \times n}$  is square and nonsingular, it satisfies

$$(A.20) \quad f'(x; M) = d_L f(x; M)M.$$

Interestingly, the LD-derivative obeys a sharp chain-rule.

PROPOSITION A.21 (Chain-rule). *Let  $h : X \rightarrow Y$  and  $g : Y \rightarrow \mathbb{R}^q$  be L-smooth at  $x \in X$  and  $h(x) \in Y$ , respectively. Then the function  $g \circ h$  is L-smooth at  $x \in X$ , and for any  $k \in \mathbb{N}$  and  $M \in \mathbb{R}^{n \times k}$ ,*

$$(A.21) \quad [g \circ h]'(x; M) = g'(h(x); h'(x; M)).$$

LD-derivatives give convenient calculus rules for the derivatives of min, max, abs, and other nonsmooth functions widely encountered in practice. These rules are based on *lexicographic ordering*, explaining why LD-derivatives are called "lexicographic". For given vectors  $x, y \in \mathbb{R}^n$ , we say that  $x$  is lexicographically lower than  $y$  if

$$(A.22) \quad \begin{aligned} x \prec y &\text{ if and only if } \exists j \in [n] \text{ such that } x_i = y_i \ \forall i < j \text{ and } x_j < y_j, \\ x \preceq y &\text{ if and only if } x = y \text{ or } x \prec y. \end{aligned}$$

The *lexicographic minimum function* returns the lexicographically ordered minimum of two vectors:

$$(A.23) \quad \mathbf{Lmin} : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n : (x, y) \mapsto \begin{cases} x & \text{if } x \preceq y, \\ y & \text{if } x \succ y. \end{cases}$$

Similarly, the *lexicographic matrix minimum function* compares two matrices row by row, and is defined as

$$(A.24) \quad \mathbf{LMmin} : \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{m \times n} : (X, Y) \mapsto \begin{bmatrix} \mathbf{Lmin}(X_1^\top, Y_1^\top) \\ \mathbf{Lmin}(X_2^\top, Y_2^\top) \\ \vdots \\ \mathbf{Lmin}(X_m^\top, Y_m^\top) \end{bmatrix}.$$

**Appendix B. Multivalued mapping.** A multivalued mapping (or point-to-set map) is a mapping  $\Gamma : E \rightrightarrows F$  where for each  $x \in E$ ,  $\Gamma(x)$  is a subset of  $F$ . They have been used extensively in mathematical programming to study the solution maps of optimization problems.

DEFINITION B.1. *Let  $\Gamma : E \rightrightarrows F$  a point-to-set map.*

- $\Gamma$  is upper semicontinuous at  $x \in E$  if for each open set  $\Omega \subset F$  satisfying  $\Gamma(x) \subset \Omega$  there exists a neighborhood  $N(x)$  of  $x$  such that for all  $y \in N(x)$ ,  $\Gamma(y) \subset \Omega$ .
- $\Gamma$  is lower semicontinuous at  $x \in E$  if for each open set  $\Omega \subset F$  satisfying  $\Gamma(x) \cap \Omega \neq \emptyset$  there exists a neighborhood  $N(x)$  of  $x$  such that for all  $y \in N(x)$ ,  $\Gamma(y) \cap \Omega \neq \emptyset$ .
- $\Gamma$  is continuous at  $x$  if it is lower semicontinuous and upper semicontinuous at  $x$ .

DEFINITION B.2. *The point-to-set map  $\Gamma : E \rightrightarrows F$  is closed at  $x$  if there exists a sequence  $x_n \in E$ ,  $x_n \rightarrow x$  such that  $y_n \in \Gamma(x_n)$  and  $y_n \rightarrow y$  imply  $y \in \Gamma(x)$ .*

DEFINITION B.3. *The point-to-set map  $\Gamma : E \rightrightarrows F$  is open at  $x$  if  $y \in \Gamma(x)$  implies there exists a sequence  $x_n \in E$ ,  $x_n \rightarrow x$  such that there exists  $m$  and  $\{y_n\}$  such that  $y_n \in \Gamma(x_n)$  for all  $n \geq m$  and  $y_n \rightarrow y$ .*

DEFINITION B.4. *The point-to-set map  $\Gamma : E \rightrightarrows F$  is uniformly compact near  $\hat{x}$  if the set  $\cup_{x \in N(\hat{x})} \Gamma(x)$  is bounded for some neighborhood  $N(\hat{x})$  of  $\hat{x}$ .*

If  $\Gamma$  is uniformly compact near  $\hat{x}$ , then  $\Gamma$  is closed if and only if  $\Gamma(\hat{x})$  is a compact set and  $\Gamma$  is upper semicontinuous at  $\hat{x}$ .

DEFINITION B.5. *The point-to-set map  $\Gamma : E \rightrightarrows F$  is upper Lipschitzian with modulus  $L$  at  $\hat{x} \in E$  if there is a neighborhood  $N$  of  $\hat{x}$  such that for each  $x \in N$ ,  $\Gamma(x) \subset \Gamma(\hat{x}) + L\|x - \hat{x}\|B$ , with  $B$  the unit-ball in  $F$ .*

**Appendix C. A primer on generalized equations.** Let  $X$  be a linear space, and  $C$  be a non-empty closed convex set in  $X$ .

DEFINITION C.1 (Normal cone). *We define the normal cone operator associated to the set  $C$  as, for  $x \in X$*

$$(C.1) \quad N_C(x) = \begin{cases} \{y \in X^* : \langle y, v - x \rangle \leq 0, \forall v \in C\} & \text{if } x \in C, \\ \emptyset, & \text{if } x \notin C. \end{cases}$$

With the normal cone, we can define formally the notion of *variational inequality*, generalizing optimization problem.

DEFINITION C.2 (Variational inequality). *Let  $p \in P$  be a parameter. For  $f : X \times P \rightarrow X^*$ , we define the parameterized variational inequality (also called generalized equation) the problem of finding  $x \in X$  such that*

$$(C.2) \quad f(x, p) + N_C(x) \ni 0.$$

The equation (C.2) is a compact formulation stating that  $-f(x, p) \in N_C(x)$ , or, equivalently

$$(C.3) \quad \langle f(x, p), v - x \rangle \geq 0, \quad \forall v \in C.$$

The solution mapping associated to (C.2) is defined as

$$(C.4) \quad S(p) = \{x \in X : f(x, p) + N_C(x) \ni 0\}.$$

Robinson extended the Implicit function theorem to the nonsmooth variational inequality (C.2), by extending the notion of *regular solution* to the variational inequality setting, where the solution mapping  $S(p)$  is not necessarily single-valued. Robinson introduced the notion of *strong regularity*, which extends the non-singularity condition we impose in the implicit function theorem. If we suppose that  $f$  is (Fréchet) differentiable, the idea is to look at the linearized generalized equation around a solution  $x_0$

$$(C.5) \quad T(x, p) \ni 0,$$

where we have defined the linearized operator at  $x_0$ :

$$(C.6) \quad T(x, p) := f(x_0, p) + \nabla_x f(x_0, p)(x - x_0) + N_C(x).$$

**DEFINITION C.3** (Strong regularity [45]). *We say that (C.2) is strongly regular at a solution  $x_0$  with Lipschitz constant  $L$  if there exists a neighborhood  $U \subset X^*$  of the origin and  $V \subset X$  of  $x_0$  such that the restriction to  $U$  of  $T^{-1} \cap V$  is a single-valued function from  $U$  to  $V$  and is  $L$ -Lipschitzian on  $U$ .*

Putting it more simply, strong regularity implies that the inverse of the linearized operator  $T^{-1}$  has a Lipschitz continuous *single-valued* localization at 0 for  $x_0$ . If  $C$  is the whole-space  $X$ , then strong regularity implies that the Jacobian  $\nabla_x f(x_0, p)$  is non-singular. The following theorem extends the implicit function theorem to generalized equations (C.2) satisfying the strong regularity condition C.3.

**THEOREM C.4** (Theorem 2.1, [45]). *Let  $p \in P$  be a given parameter,  $x_0$  solution of the generalized equation (C.2). Suppose that  $f$  is Fréchet-differentiable w.r.t.  $x$  exists at  $p$ , and that both  $f(\cdot, \cdot)$  and  $\nabla_x f(\cdot, \cdot)$  are continuous at  $(x_0, p)$ . If (C.2) is strongly regular at  $x_0$ , then for any  $\varepsilon > 0$  there exists neighborhood  $N$  of  $p$  and  $W$  of  $x_0$  as well as a single-valued function  $x : N \rightarrow W$  such that for any  $p \in N$ ,  $x(p)$  is the unique solution of the inclusion*

$$(C.7) \quad f(x, p) + N_C(x) \ni 0,$$

and, for all  $p, q \in N$ , one has

$$(C.8) \quad \|x(p) - x(q)\| \leq (L + \varepsilon) \|f(x(p), p) - f(x(q), q)\|.$$

Theorem C.4 states that the strong regularity of (C.2) implies that the solution mapping  $S(p)$  of (C.2) has a *Lipschitz-continuous* single-valued localization at  $p$  for  $x_0$ .

**COROLLARY C.5** (Locally-Lipschitz condition). *Suppose the hypotheses of Theorem C.4 hold. Suppose in addition there exists a constant  $\kappa$  such that for each  $(p, q) \in N$ , for each  $x \in W$ , one has*

$$(C.9) \quad \|f(x, p) - f(x, q)\| \leq \kappa \|p - q\|.$$

*Then  $x(\cdot)$  is Lipschitzian on  $N$  with modulus  $\kappa(L + \varepsilon)$ .*